

100

i	Query	Score	Match Length	DB	Description		
1	2056	58	9	390	4	060667	homo sapien

2	1194.5	34.2	11	Q9D811	Q9D811 mus musculus	
3	202.5	5.8	455	11	Q92038	Q92038 mus musculus
4	200.5	5.8	535	11	Q31037	Q31037 mus musculus
5	202.5	5.7	532	4	Q8M4V6	Q8M4V6 homo sapien
6	198.5	5.7	534	4	Q6S8A2	Q6S8A2 homo sapien
7	192	5.5	731	6	Q8S16	Q8S16 mus musculus
8	187.5	5.4	1175	11	P70200	P70200 mus musculus
9	186	5.3	1175	11	Q35126	Q35126 mus musculus
10	184	5.3	758	6	Q9N2H7	Q9N2H7 sus scrofa
11	183	5.2	733	6	Q8S083	Q8S083 trichosturus
12	179	5.1	616	4	Q9H6K5	Q9H6K5 homo sapien
13	178	5.1	1182	4	Q99495	Q99495 homo sapien
14	178	5.1	1180	4	Q99621	Q99621 homo sapien
15	175.5	5.0	1008	11	Q9DBV8	Q9DBV8 mus musculus
16	175.5	5.0	1734	11	Q60592	Q60592 mus musculus
17	173.5	5.0	995	4	Q9Y2M4	Q9Y2M4 homo sapien
18	173	5.0	950	5	Q9V6A5	Q9V6A5 mus musculus
19	173	5.0	4957	4	Q14687	Q14687 homo sapien
20	172.5	4.9	1012	4	Q43393	Q43393 homo sapien
21	171.5	4.9	1012	4	Q75353	Q75353 homo sapien
22	170.5	4.9	305	11	Q8V1M2	Q8V1M2 mus musculus
23	168.5	4.8	939	16	Q8KH40	Q8KH40 talstonia s
24	167	4.8	680	12	Q9Q0C4	Q9Q0C4 chlayota mosca
25	166	4.8	448	12	Q9J2B0	Q9J2B0 maraca mula
26	165.5	4.7	448	12	Q9J259	Q9J259 maraca mula
27	165.5	4.7	1400	4	Q9HPS7	Q9HPS7 mus sapien
28	164	4.7	1404	12	Q9Q0F3	Q9Q0F3 ceropithec
29	164	4.7	1175	6	Q9V1V6	Q9V1V6 ceropithec
30	164	4.7	2703	5	Q9VHG7	Q9VHG7 oryctolagus
31	163	4.7	1724	5	P91019	P91019 drosophila
32	162	4.6	336	10	Q22514	Q22514 caenorhabdi
33	162	4.6	2715	5	Q61603	Q61603 santalum al
34	161	4.6	307	11	Q54947	Q54947 drosophila
35	161	4.6	448	12	Q9MR42	Q9MR42 tatus norveg
36	160	4.6	1315	10	Q9SRM0	Q9SRM0 maraca mula
37	159.5	4.6	1565	4	Q9P2P6	Q9P2P6 zea mays (tr
38	159.5	4.6	2158	11	Q9PMU13	Q9PMU13 ratius norveg
39	159	4.6	605	12	Q9Q5L3	Q9Q5L3 ceropithec
40	159	4.6	1044	4	Q9F206	Q9F206 homo sapien
41	159	4.6	1234	3	Q13637	Q13637 schizosarc
42	159	4.6	1234	3	Q9F6T1	Q9F6T1 leucospora
43	158.5	4.5	1059	4	Q8TCC7	Q8TCC7 homo sapien
44	158.5	4.5	2167	11	Q9HVV48	Q9HVV48 ratius norveg
45	157.5	4.5	939	6	Q9G1W0	Q9G1W0 canis fami

RL Submitted (Apr-2021) to the EMBL/GenBank/DBJ databases  
 DR EMBL: AF057557; AAC18930.1; -  
 DR EMBL: BC006401; AA006401.1; -  
 DR InterPro: IPR003599; Iq  
 DR InterPro: IPR003006; Tq\_MHC  
 Pfam: PF00047; Iq\_1  
 DR SMART: SM00409; Iq\_1  
 SO SOURCE: 390 AA; 43146 MW; F1931D217EHCAG96C CNG6;

Alignment Scores:  
Pred No.

Pred. No	4 096 168	Length	396
Score:	2056.00	Matches:	396
Percent Similarity:	100.008	Conservative:	0
Best local Similarity:	100.008	Mismatches:	0
Query Match:	58.91%	Indels:	0
DB:	4	Gaps:	0

US-09-651-1508-1 (1-1911) x 060667 (1-390)

74 ATGCACCTTCCTGGCTTGGCCACTTTCCTGCGACGATACAGGGCCCTCAGGATCCCTC 133  
 1 MetAspPheTrpLeuTrpProLeuTrpPheLeuProValSerGlyAlaLeuArgIleLeu 20

QY 134 CCGAGACTAAAGCTAGAGGGGAGCGCTGGGGCGAGTACCATCAATGCCACCTTCT  
 |||||  
 Db 21 ProCluVal<sub>1</sub>ysVal<sub>2</sub>Gluc<sub>3</sub>ylcIu<sub>4</sub>rcu<sub>5</sub>ylcyl<sub>6</sub>ys<sub>7</sub>er<sub>8</sub>Val<sub>9</sub>Thr<sub>10</sub>Iel<sub>11</sub>ys<sub>12</sub>Cys<sub>13</sub>Pro<sub>14</sub>Leu<sub>15</sub>Pro<sub>16</sub>

Dy 194 GAAATGCATGTGAAGATAATCTGTAACCAAGA;ATGGTTCATATCGAACAATGTGTAAC  
 ++++++  
 Db 41 GUMETHISVALArgIleIyrIleUCysArgGIumETalacIylserGIyThCysGIyThh

QY 254 GTCGTAATCCACCAACCTTCATTCAGGCGAGAAATACAGAGCGCCGAGTTACTCTTCACCAACA  
|||||  
Db 61 ValValSerThrThrAsnHeiLeuValagIuTVrLysGLVaRqValThrLeuLysGln

QY 314 TACCCACGCAAGATCTGTCTTCTACTTCGAGGTAAACACAGCTGACAGAAAGTACAGCCGCGG  
|||||  
Db 81 TTTTATGATGVSASnleupheleuValGluValTPrCnleuphTnCnleuphTnCSnASnSCTGle

[illegible]

434 AATGTCACAGTGAATACGAGCCATCATGCGAAGSAGCAGCCAAATGGCTTGAGACCTCGAAAA

494 TTTCTTTCATCTGGGCTATTGTGTTCCATAATGCGCTTCGCGCAAA  
|||||

Qy 554 ACCAGGATTACAACTAAGTAAAAGGGAAAGGTCTTTCAATTCACAAACTCCTCCCG

614 ACCATCCCAATCACCACCGGCTGCACTGCTGAGAGTATCTTCACCTAGCTAGCTGCAGAA

DB 181 ThrThrGlnIleThrHisArgProAlaValSerArgAlaSerSerValAlaGlyAspLys  
QY 674 CCGCGAAGCTTTCCTGGCATTCACACTACAGCCTCAAAAATCTCAGCTCTCGAAGGCCCTGCATC

Db 201 P T G A G T G T T h e l e u p r o S e r T h r T h r A l a S e r T y s I l e S e r A l a L e u C h u C l y l e L e u C

Db 221 TysTtOClInThrProSerTytAsnHisHisIleAlaGluHisArgGlnAlaGalaLeuAsp

Db 241 TyrGlySerGlnSerGlyArgGlnGlnGlyGlnGlyPheHisIleLeuIleProThrIleLeu

[illegible]

RESULT 2  
09D8T1

ID	Q9D8T1	PRELIMINARY;	PRT;	422	AA
AC	0000001				

DT 01-JUN-2001 (TREMHAIRE), 17, Created

DT 01-JUN-2002 (TREMBL) 21, Last annotation update  
DE 1810037R05Pik protein

GN	18100037B05PRK
OS	MUS MUSCULUS (Mouse)

OC Eukaryota, Metazoa  
OC Mammalia, Mammalia

OX	NCBI_TaxID=10090,
BN	[1]

SEQUENCE FROM N.2

RX	MEDLINE: 21085660
RA	Kawai J, Shinaga

KA Arakawa T., Hara  
KA Aizawa K., Iizawa

RA Saito T., (Kazak  
Kadota K., Matsui

KA Fleischmann W., O  
KA Kuehl P., Lewis S

KA SCHRIM L.M., Ste  
KA Saka K., Okido T

KA Blake J., Bolton  
KA Brownstein M.J.,

RA JUSTINICH S., H  
RA LYONS P., Marchig

RA Sasaki H., Sato K

WYLSHAW-RUTIS A

KT "functional; annotated"

EMBI, AK007714, I

InterPro; IPR0036

SEQUENCE 422 A

Alignment Scores:

Secure:

**SECRET**



















"NEED human cDNA sequencing project."  
 BL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK025437; BAB15254.1.  
 DR InterPro: IPR002965; P\_Rich\_extensn.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 SV SOURCE: 516 AA; 60038 MW; 584414376RCER; CRC64;

## Alignment Scores:

Score:	4.55e-06	Length:	616
Percent Similarity:	179.00	Matches:	156
Best local Similarity:	32.798	Conservative:	45
Query Match:	25.458	Mismatches:	220
Indels:	5.138	Indels:	192
Gaps:	4	Gaps:	31

US-09-651-150b-1 (1-1911) x Q9H6K5 (1-616)

```

OY 286 ATCAAGAGCCGCTACTGCAAGCAATACCCAGACCAATCTGCTACTGAGCT 345
DB 15 LeuGlnThrProAlaSerLeuThrProLeuGlnLeuProSerLeuAla 34
OY 346 -----AACACAGCTGATATAAGTATATAGTTATAGTTATAGTATAGT 393
DB 35 ProProProLeuGlnThrAla-----SerAlaProLeuThrPro 48
OY 394 CATGACAGCAAGCCGCGAAGAACCCAGAAAGTCACCTGATGTCACAGTATACGA 453
DB 49 HisLeuGlnThrPro-----ProCysProAlaPro-----CysPro-LeuGlnAlaPro 64
OY 454 GCGATCATGAGACAGACAGCAATGCTGAGACTGCCAAATGCTTCACTGCGCTATT 513
DB 64 GProSerProLeuThrThrProProProGlnThrProSerSerLeuAlaThrProPro 84
OY 514 GTTCAGATATGATATATATATATATATATATATATATATATATATATATAT 555
DB 84 GAlaAlaAlaProLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 104
OY 556 CAGAGTATACCAAGCAAGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 603
DB 104 ProLeuAlaThrProProProGlnAlaProProAlaLeuAlaLeuProLeuGlnA 124
OY 604 CTGCGCCGCGCGAGCCGCAAA-----TCACCCAGCGCGCGCGAGTGTCCAGAG 650
DB 124 LProProSerProProAlaSerProProLeuSerProLeuAlaThrProSerProGln 144
OY 651 -----CATCTCAGTATAGTATGATGATGATGATGATGATGATGATGAT 690
DB 144 LProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 163
OY 691 ATCCACTATACAG-----CCTCAAAAATCTCAGCTCTGCA 723
DB 163 GProValAlaGlnAlaProPheserProProAlaSerProProValAlaSerPro-SerAlaThr 182
OY 724 GCGGCTGCTCAAGCCGCGAGAGCCAGCTACAAACAGCCAGCCAGCGCTGCGAGAG 783
DB 183 ProProSerGlnAlaAlaProProSerLeuAlaAlaProProLeuGlnAlaThrProSerPro 202
OY 784 AGCACAGTATATATATATATATATATATATATATATATATATATATAT 834
DB 203 ProAlaSer-----ProProMetSerProSerAlaThr 213
OY 835 CCTATGCGCGAGCATCTGCGGCTTT-----CCT 864
DB 214 ProProProGlnAlaProProProLeuAlaAlaProProLeuGlnAlaProProSerPro 233
OY 865 CCTGCTACTTCTGCGGCTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 924
DB 234 ProAlaSerProProMetSerProSerAlaAlaThrProProAlaAlaThrProProLeu 253
OY 925 GCGAGTATATATATATATATATATATATATATATATATATATATATATATAT 984
DB 254 AlAlaAlaProProLeuGlnAlaThrProProSerProProAlaSerProProMetSerPro 273

```

```

OY 985 GTC-----GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1029
DB 274 AlAlaSerProProGlnAlaAlaProProAlaLeuAlaThrProProLeuGlnAlaAlaPro 293
OY 1030 GCGAGTATATATATATATATATATATATATATATATATATATATATATATAT 1089
DB 294 Ser-----Pro-ProAlaSerProProGlnAlaAlaProProSerPro 308
OY 1090 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1140
DB 308 LAlaSerLeuProMetSerProLeuAlaThrProProProGlnAlaAlaProProValAla 328
OY 1141 TCGGATCTCTCAACAGCAAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 328 AlAlaProLeuGlnAlaAlaProProSer-----ProProAlaSer 341
OY 1201 GATGATATATATATATATATATATATATATATATATATATATATATATATATAT 1260
DB 342 -----ProThrLeuAlaAlaProAlaThr 349
OY 1261 TCGGATCTCTCAACAGCAAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 349 GProProThrProProProProAlaThrSerValSerGlyProAlaLeuThrLeuAlaLeu 368
OY 1321 ATACCTGCTTCAATGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 369 -----AlaProGlyProProProPro-ProProSerAlaAlaSerPro 380
OY 1381 TCGGATATATATATATATATATATATATATATATATATATATATATATATATAT 1425
DB 381 SerSerThrLeuSerGlyProAlaLeuAlaGlyHisSerSerAlaAlaThrSerPro 400
OY 1425 ----- 1425
DB 401 GLeuLeuAlaGlyGlyTyrAspSerGlyProSerGlyGlyAlaAlaAlaAlaSerProProPro 420
OY 1426 -----TCCTGACAGCT-----TGCGAG-----CCGCGTGA 1452
DB 421 AspAlaGlyLeuAlaAlaAlaGlyHisProAlaAlaThrSerAlaGlyProAlaProProPro 440
OY 1453 GTTACAGAGTATATATATATATATATATATATATATATATATATATATATATATAT 1512
DB 441 AlAlaAlaGlyAlaProGlyAlaAlaProLeuProThrProProAlaThr 456
OY 1513 AAGCGTATGAT-----AGATATATATATATATATATATATATATATATATATATAT 1557
DB 457 GlyProGlySerAlaAlaAlaGlyLeuGlySerThrLeuThrLeuGlyProAlaSerAla 476
OY 1558 TCTTTGACCTACAGAGTGTGTTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596
DB 477 ThrProAlaProGly-----AlaLeuSerProGlyProSerProGlyThr 491
OY 1597 -----CATATGATATATATATATATATATATATATATATATATATATATATAT 1647
DB 492 SerGlyGlyHisAlaAlaAlaGlyAlaGlyAlaAlaAlaSerSerAlaSerProGlyGln 511
OY 1648 GATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1668
DB 512 AlAlaGlyLeuGlyGlyLeuProLeuGlyAlaLeuAlaAlaSerValAlaGlnHisLeuLeu 531
OY 1669 AATACAGAGCTTCTGCTGCTGCTATATGAGTGTGCTGCGAGAGAGAGAGAGAGAGAGAG 1728
DB 532 SerAlaGlnThrLeuLeu-----AlaAlaAlaGlyAlaAlaAlaGlyGly 546
OY 1729 TCTGATATATATATATATATATATATATATATATATATATATATATATATATATAT 1755
DB 547 SerGly-----GlyGlyProGlyGlyAla 554

```

RESULT 13  
 G99495  
 ID G99495 PRELIMINARY: PFT: 1182 AA.  
 AC G99495.  
 DT 01-MAY-1997 (TREMblrel. 03, Created)







## RESULT 15

Q9DBV8

ID Q9DBV8 PRELIMINARY: PRT, 1008 AA.

DT 01-JUN-2001 (TREMUR, 17, Created)

DT 01-JUN-2001 (TREMUR, 17, last sequence update)

DT 01-JUN-2002 (TREMUR, 21, last annotation update)

DE Microtubule associated lesion specific serine/threonine protein kinase.

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

Q9DBV8

ID Q9DBV8 PRELIMINARY: PRT, 1008 AA.

DT 01-JUN-2001 (TREMUR, 17, Created)

DT 01-JUN-2001 (TREMUR, 17, last sequence update)

DT 01-JUN-2002 (TREMUR, 21, last annotation update)

DE Microtubule associated lesion specific serine/threonine protein kinase.

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

Q9DBV8

ID Q9DBV8 PRELIMINARY: PRT, 1008 AA.

DT 01-JUN-2001 (TREMUR, 17, Created)

DT 01-JUN-2001 (TREMUR, 17, last sequence update)

DT 01-JUN-2002 (TREMUR, 21, last annotation update)

DE Microtubule associated lesion specific serine/threonine protein kinase.

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

Q9DBV8

ID Q9DBV8 PRELIMINARY: PRT, 1008 AA.

DT 01-JUN-2001 (TREMUR, 17, Created)

DT 01-JUN-2001 (TREMUR, 17, last sequence update)

DT 01-JUN-2002 (TREMUR, 21, last annotation update)

DE Microtubule associated lesion specific serine/threonine protein kinase.

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

Q9DBV8

ID Q9DBV8 PRELIMINARY: PRT, 1008 AA.

DT 01-JUN-2001 (TREMUR, 17, Created)

DT 01-JUN-2001 (TREMUR, 17, last sequence update)

DT 01-JUN-2002 (TREMUR, 21, last annotation update)

DE Microtubule associated lesion specific serine/threonine protein kinase.

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

Q9DBV8

ID Q9DBV8 PRELIMINARY: PRT, 1008 AA.

DT 01-JUN-2001 (TREMUR, 17, Created)

DT 01-JUN-2001 (TREMUR, 17, last sequence update)

DT 01-JUN-2002 (TREMUR, 21, last annotation update)

DE Microtubule associated lesion specific serine/threonine protein kinase.

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

DE

```
Db 805 ThrAsnProLeuGlnValProSerLeuSerArg-----SerGlyPro 818
QY 1489 CCGAATGGCGACTTGGCTTCCCTTCCAAAGCCCTGGACAGACTGTGGATTTCAGAGACTT 1548
      ::: ||| ||||| ||| |||
Db 819 ThrSerProThr-----ProSerGluClyCysTrpLysAlaGlnHisLeu 833
QY 1549 TCTTCATGTCTTTGAGC-----ACAGGGTGT 1575
      ::: ||||| |||||
Db 834 HisThrGlnAlaLeuThrAlaLeuGlyProSerPheSerGluLeuThrProThrClyCys 853
QY 1576 TGTTCGTCGCAAGCTCTACATCAGATGGCAATCAGGCTGC----- 1614
      ::: |||
Db 854 SerAlaAlaThrSerThrSerGlyLysProGlyThrTrpSerTrpLysPheLeuIleGlu 873
QY 1615 GGCAGAGCGCAATATATATTCCTGGGAGATCCCTGGCAGAGCTTGGCTTACACAAATAGA 1674
      ||| ||| ::: ||| |||
Db 874 GlyProAspArgAlaSerThrAsnLysIleMetThrArgLysClyGluProAlaAsnSer 893
QY 1675 AGG-----CTCTGCTCTGAGTATGT-----GACGT 1701
      ::: ||||| |||
Db 894 GlnAspThrAsnThrThrValProAsnIleLeuLysAsnIleuSerProGluClyGly 913
QY 1702 GCCTCAGCCCCAGACTAAGCAGAGGCTCTGATATAACACTCTCGAAACGGCT 1756
      ||||| ||| ::: ||||| |||
Db 913 sProGlnProProSerValPro---GlyLeuThrHisProLeuGlnValPro 930
```

Search completed: November 24, 2002, 05:58:32  
Job time : 105.5 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein protein search using SW model-1

Run on: November 24, 2002, 06:11:16, Search time: 40 seconds  
(without alignments)  
573.747 Million cell updates/sec

Title: US-09-651-150B-2  
Perfect score: 2055  
Sequence: 1 MDRWLEPLFLPLVSGALPIL ..... HGPAMMETSPDELYINVA 390

Scoring table: HUSUM62  
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents: AA: \*  
1: /cgn2\_6/producta/2/1aa/5A.COMR pep: \*  
2: /cgn2\_6/producta/2/1aa/5B.COMR pep: \*  
3: /cgn2\_6/producta/2/1aa/5A.COMR pep: \*  
4: /cgn2\_6/producta/2/1aa/5A.COMR pep: \*  
5: /cgn2\_6/producta/2/1aa/5A.COMR pep: \*  
6: /cgn2\_6/producta/2/1aa/5A.COMR pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	2047	99.6	390	3	US-09-961-564A-2
2	1157	56.3	422	4	US-09-724-864-45
3	550	26.8	107	3	US-08-961-564A-4
4	198	9.6	771	3	US-08-434-000A-8
5	198	9.6	771	4	US-09-312-157-8
6	187	9.1	109	3	US-08-961-564A-9
7	185	9.0	769	3	US-08-434-000A-10
8	185	9.0	769	4	US-09-312-157-10
9	182	8.9	608	4	US-09-095-385-4
10	182	8.9	746	3	US-08-434-000A-4
11	182	8.9	746	4	US-09-312-157-4
12	182	8.9	757	3	US-08-434-000A-6
13	182	8.9	757	4	US-09-312-157-6
14	166.5	8.1	624	2	US-08-643-406A-2
15	166.5	8.1	624	4	US-09-199-534-22
16	166.5	8.1	624	4	US-09-199-534-22
17	166.5	8.1	773	3	US-08-434-000A-2
18	166.5	8.1	773	4	US-09-312-157-2
19	117.5	5.7	476	3	US-08-487-550-4
20	117	5.7	453	3	US-08-466-151-8
21	117	5.7	453	4	US-08-466-151-8
22	113.5	5.5	2337	3	US-08-713-118-2
23	113.5	5.5	2337	4	US-09-452-007-2
24	112	5.5	451	2	US-08-887-352H-14
25	112	5.5	451	3	US-08-887-352H-16
26	112	5.5	451	3	US-08-466-151-65
27	112	5.5	451	4	US-09-109-207C-14

28	112	5.5	451	4	US-09-109-207C-16	Sequence 16, Appl
29	112	5.5	451	4	US-09-296-005-14	Sequence 14, Appl
30	112	5.5	451	4	US-09-296-005-16	Sequence 16, Appl
31	111	5.4	476	2	US-08-378-939-10	Sequence 10, Appl
32	109.5	5.3	2339	1	US-08-455-543A-47	Sequence 47, Appl
33	109.5	5.3	2339	2	US-08-455-543A-47	Sequence 47, Appl
34	109.5	5.3	2339	4	US-09-268-163-6	Sequence 6, Appl
35	109.5	5.3	2339	4	US-09-268-163-6	Sequence 6, Appl
36	107	5.2	292	4	US-09-345-468-18	Sequence 18, Appl
37	107	5.2	292	4	US-09-345-468-18	Sequence 18, Appl
38	107	5.2	313	4	US-09-345-468-16	Sequence 16, Appl
39	107	5.2	313	4	US-09-414-453A-16	Sequence 16, Appl
40	107	5.2	449	1	US-08-458-516-13	Sequence 13, Appl
41	106	5.2	390	2	US-08-979-424-1	Sequence 1, Appl
42	105	5.1	451	2	US-08-887-352H-18	Sequence 18, Appl
43	105	5.1	451	4	US-09-109-207C-18	Sequence 18, Appl
44	105	5.1	451	4	US-09-282-505-2	Sequence 2, Appl
45	105	5.1	451	4	US-09-054-255-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-961-564A-2  
Sequence 2, Application US/08961564A

Patent No. 6114515

GENERAL INFORMATION:

APPLICANT: WU, SHOUJIAN

APPLICANT: SWEET, RAYMOND

APPLICANT: TRINER, ALEMESEGED

TITLE OF INVENTION: PIGRL-1, A MEMBER OF IMMUNOGLOBULIN

NUMBER OF SEQUENCES: 9

COMMUNICATION ADDRESS:

ADDRESSEE: RATNER & PRESTIA

STREET: P.O. BOX 980

CITY: VALLEY FORGE

STATE: PA

COUNTRY: USA

ZIP: 19482

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSD for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,564A

FILING DATE: 30-OCT-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/056,935

FILING DATE: 25-AUG-1997

ATTORNEY/AGENT INFORMATION:

NAME: PRESTIA, PAUL F

REGISTRATION NUMBER: 23,031

REFERENCE/DOCKET NUMBER: CH-70236

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-407-0700

TELEFAX: 610-407-0701

TELEX: 846169

INFORMATION FOR SEQ ID NO. 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 390 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MODEL TYPE: protein

US-08-961-564A-2

Query Match: 99.6%, Score 2047, DB 3, Length 390;

Post Local Similarity: 99.7%, Pred No. 2, 3e 178;

Matches 389, Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MDRWMLPFLPVSGALRIPEVVEVNGISGVTIKPILPEMHVLYCEEMAGSGTGT 60
  1 MDRWMLPFLPVSGALRIPEVVEVNGISGVTIKPILPEMHVLYCEEMAGSGTGT 60
QY 61 VVSTNFTKAEYKGVTLKQYPRKNLPLVEVYQLTESDSGVYAGAMNTDRTKQAVTL 120
  61 VVSTNFTKAEYKGVTLKQYPRKNLPLVEVYQLTESDSGVYAGAMNTDRTKQAVTL 120
QY 121 NVHSFYSMEQPMPEPPEKMFHLYPLFQMPAYASSSKFVTRVTPQOKKVPVHHSSP 180
  121 NVHSFYSMEQPMPEPPEKMFHLYPLFQMPAYASSSKFVTRVTPQOKKVPVHHSSP 180
QY 181 TTQITHRRVRSASSVAGDKPRTFLPSTTAKISALSKLLKQTPSYNHHHTPLHROPALD 240
  181 TTQITHRRVRSASSVAGDKPRTFLPSTTAKISALSKLLKQTPSYNHHHTPLHROPALD 240
QY 241 VVSQSGREGQGFHLLPILGLFLALLGLVYKRAVERRRALSRARLAVRRALAESQ 300
  241 VVSQSGREGQGFHLLPILGLFLALLGLVYKRAVERRRALSRARLAVRRALAESQ 300
QY 301 RPKGSPRPSQNNIYSACPRKAKADAAAGCEAPVPGGAPLPAPLQVSESPWLIHAPSL 360
  301 RPKGSPRPSQNNIYSACPRKAKADAAAGCEAPVPGGAPLPAPLQVSESPWLIHAPSL 360
QY 361 KTSCEYVSLYHQPAAMMEDSDSDYINVA 390
  361 KTSCEYVSLYHQPAAMMEDSDSDYINVA 390

```

## RESULT 2

```

US-09-724-864-45
: Sequence 45, Application US/09724864
: Patent No. 6380362
: GENERAL INFORMATION:
: APPLICANT: Watson, James D.
: APPLICANT: Murison, James G.
: TITLE OF INVENTION: Polynucleotides, polypeptides expressed
: TITLE OF INVENTION: by the polynucleotides and methods for their use.
: FILE REFERENCE: 11000.105001
: CURRENT APPLICATION NUMBER: US/09/724, 864
: PRIORITY FILING DATE: 2000-11-28
: PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
: PRIORITY FILING DATE: 1999-12-23
: NUMBER OF SEQ ID NOS: 72
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 45
: LENGTH: 422
: TYPE: PRT
: ORGANISM: Mouse
US-09-724-864-45

```

Query Match 56.38, Score 1157, DB 4, Length 422;

Best Local Similarity 57.58; Pred. No. 2.8e-97;

Matches 234; Conservative 46; Mismatches 107; Indels 20; Gaps 6;

```

QY 1 MDRWMLPFLPVSGALRIPEVVEVNGISGVTIKPILPEMHVLYCEEMAGSGTGT 60
  1 MDRWMLPFLPVSGALRIPEVVEVNGISGVTIKPILPEMHVLYCEEMAGSGTGT 60
QY 61 VVSTNFTKAEYKGVTLKQYPRKNLPLVEVYQLTESDSGVYAGAMNTDRTKQAVTL 120
  61 VVSTNFTKAEYKGVTLKQYPRKNLPLVEVYQLTESDSGVYAGAMNTDRTKQAVTL 120
QY 121 NVHSFYSMEQPMPEPPEKMFHLYPLFQMPAYASSSKFVTRVTPQOKKVPVHHSSP 180
  121 NVHSFYSMEQPMPEPPEKMFHLYPLFQMPAYASSSKFVTRVTPQOKKVPVHHSSP 180
QY 181 TTQITHRRVRSASSVAGDKPRTFLPSTTAKISALSKLLKQTPSYNHHHTPLHROPALD 240
  181 TTQITHRRVRSASSVAGDKPRTFLPSTTAKISALSKLLKQTPSYNHHHTPLHROPALD 240
QY 241 VVSQSGREGQGFHLLPILGLFLALLGLVYKRAVERRRALSRARLAVRRALAESQ 300
  241 VVSQSGREGQGFHLLPILGLFLALLGLVYKRAVERRRALSRARLAVRRALAESQ 300
QY 301 RPKGSPRPSQNNIYSACPRKAKADAAAGCEAPVPGGAPLPAPLQVSESPWLIHAPSL 360
  301 RPKGSPRPSQNNIYSACPRKAKADAAAGCEAPVPGGAPLPAPLQVSESPWLIHAPSL 360
QY 361 KTSCEYVSLYHQPAAMMEDSDSDYINVA 390
  361 KTSCEYVSLYHQPAAMMEDSDSDYINVA 390

```

```

1b 234 EAPTHHSHPHVGFEDHSHHPIPEFHLLPFLFLVLLSLVKKALQJPRFASSTKFAJ 298
  234 EAPTHHSHPHVGFEDHSHHPIPEFHLLPFLFLVLLSLVKKALQJPRFASSTKFAJ 298
QY 288 RLAVMKRLAES-----QRPVSPRPSQNNIYSACPRKAKADAAAGCEAPVPGGAPLPAPLQVSESPWLIHAPSL 342
  288 RLAVMKRLAES-----QRPVSPRPSQNNIYSACPRKAKADAAAGCEAPVPGGAPLPAPLQVSESPWLIHAPSL 342
QY 343 PPAVLQVSESPWLIHAPSLKTSCEYVSLYHQPAAMMEDSDSDYINVA 389
  343 PPAVLQVSESPWLIHAPSLKTSCEYVSLYHQPAAMMEDSDSDYINVA 389
QY 359 SHASHQVLLAPVPHHPIPSIKMSCEYVSLYHQPAAMMEDSDSDYINVA 405
  359 SHASHQVLLAPVPHHPIPSIKMSCEYVSLYHQPAAMMEDSDSDYINVA 405

```

## RESULT 3

```

US-08-961-564A-4
: Sequence 4, Application US/08961564A
: Patent No. 6114515
: GENERAL INFORMATION:
: APPLICANT: WO, SHOUJIAN
: APPLICANT: SWEET, RAYMOND
: APPLICANT: TRINER, ALEMESEGED
: TITLE OF INVENTION: PIGRL-1, A MEMBER OF IMMUNOGLOBULIN
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: RATNER & PRESTIA
: STREET: P.O. BOX 940
: CITY: VALLEY FORGE
: STATE: PA
: COUNTRY: USA
: ZIP: 19482
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/961, 564A
: FILING DATE: 30-OCT-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/056, 935
: FILING DATE: 25-AUG-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: PRESTIA, PAUL F.
: REGISTRATION NUMBER: 23,031
: REFERENCE/DOCKET NUMBER: GH-70236
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-407-0700
: TELEFAX: 610-407-0701
: TELEX: 846169
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 107 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-961-564A-4

```

Query Match 26.88; Score 550; DB 3; Length 107;

Best Local Similarity 90.08; Pred. No. 7.5e-43;

Matches 104; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MDRWMLPFLPVSGALRIPEVVEVNGISGVTIKPILPEMHVLYCEEMAGSGTGT 60
  1 MDRWMLPFLPVSGALRIPEVVEVNGISGVTIKPILPEMHVLYCEEMAGSGTGT 60
QY 61 VVSTNFTKAEYKGVTLKQYPRKNLPLVEVYQLTESDSGVYAGAMNTDRTKQAVTL 120
  61 VVSTNFTKAEYKGVTLKQYPRKNLPLVEVYQLTESDSGVYAGAMNTDRTKQAVTL 120
QY 121 NVHSFYSMEQPMPEPPEKMFHLYPLFQMPAYASSSKFVTRVTPQOKKVPVHHSSP 180
  121 NVHSFYSMEQPMPEPPEKMFHLYPLFQMPAYASSSKFVTRVTPQOKKVPVHHSSP 180
QY 181 TTQITHRRVRSASSVAGDKPRTFLPSTTAKISALSKLLKQTPSYNHHHTPLHROPALD 240
  181 TTQITHRRVRSASSVAGDKPRTFLPSTTAKISALSKLLKQTPSYNHHHTPLHROPALD 240
QY 241 VVSQSGREGQGFHLLPILGLFLALLGLVYKRAVERRRALSRARLAVRRALAESQ 300
  241 VVSQSGREGQGFHLLPILGLFLALLGLVYKRAVERRRALSRARLAVRRALAESQ 300
QY 301 RPKGSPRPSQNNIYSACPRKAKADAAAGCEAPVPGGAPLPAPLQVSESPWLIHAPSL 360
  301 RPKGSPRPSQNNIYSACPRKAKADAAAGCEAPVPGGAPLPAPLQVSESPWLIHAPSL 360
QY 361 KTSCEYVSLYHQPAAMMEDSDSDYINVA 390
  361 KTSCEYVSLYHQPAAMMEDSDSDYINVA 390

```

## RESULT 4

US-08-434-000A-8  
Sequence 8, Application US/08434000A  
Patent No. 6046037  
GENERAL INFORMATION:  
APPLICANT: ANDREW C. HIATT, JULIAN  
APPLICANT: K.-C. MA, THOMAS LEHNER  
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 643 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
SOFTWARE: Word Perfect 5.1  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/434,000A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below: 1  
APPLICATION NUMBER: 08/367,395  
FILING DATE: 12/30/94  
ATTORNEY/AGENT INFORMATION:  
NAME: Guise, Jeffrey W.  
REGISTRATION NUMBER: 34,613  
REFERENCE/DOCKET NUMBER: 212/127  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 552-8400  
TELEFAX: (619) 552-0159  
TELEX: 67-3510  
SEQUENCE LISTING:  
INFORMATION FOR SEQ. ID NO.: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 771 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
DESCRIPTION: Mouse polyclonal immunoglobulin Receptor  
US-08-434-000A-8

Query Match 9.6%; Score 198, DB 3, Length 771,  
Best Local Similarity 31.5%; Pred. No. 1, 3e-09;  
Matches 53; Conservative 23; Mismatches 68; Indels 24; Gaps 5.

DB 10 FLVPSGALRIIPVYKVGCHLGGSTVTKCPHPE---MHVRIYICRFMAGSGTCTVASTT 65  
13 FSGVSTKSPITGPOEVSSIEGDSVITCYPPDTSVNHRTKRYWGRGA-SGMCTTIISSN 71  
QY 66 NIKAEYKGRVTLKQYPRKKNLFVEVTLTSTGSAVYACGAMNITGKTKYKTLNVHSE 125  
DB 72 GYLSKYSGRANLINFPRNNTFVINIHOITODDGSYKCGIG-TSNQGLSHVSLSEV--- 127  
QY 126 YEPSNEQPMPEPKMFLPYLFQMPAVASSSKFVITVTPAQRGKVP 173  
DB 128 -----SQVPELPSTHIV-----YTKDIGRNVITIECPFKRENVP 160

RESULT 5  
US-09-312-157-8  
Sequence 8, Application US/09312157  
Patent No. 6303341  
GENERAL INFORMATION:  
APPLICANT: ANDREW C. HIATT, JULIAN

K.-C. MA, THOMAS LEHNER  
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
PROTEINS IN PLANTS AND THEIR USES  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 643 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
SOFTWARE: storage  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/312,157  
FILING DATE: 14-May-1999  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/434,000  
FILING DATE: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Guise, Jeffrey W.  
REGISTRATION NUMBER: 34,613  
REFERENCE/DOCKET NUMBER: 212/127  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 552-8400  
TELEFAX: (619) 552-0159  
TELEX: 67-3510  
SEQUENCE LISTING:  
INFORMATION FOR SEQ. ID NO.: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 771 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
DESCRIPTION: Mouse Polyclonal immunoglobulin Receptor  
US-09-312-157-8

Query Match 9.6%; Score 198; DB 4; Length 771;  
Best Local Similarity 31.5%; Pred. No. 1, 3e-09;  
Matches 53; Conservative 23; Mismatches 68; Indels 24; Gaps 5;

DB 10 FLVPSGALRIIPVYKVGCHLGGSTVTKCPHPE---MHVRIYICRFMAGSGTCTVASTT 65  
13 FSGVSTKSPITGPOEVSSIEGDSVITCYPPDTSVNHRTKRYWGRGA-SGMCTTIISSN 71  
QY 66 NIKAEYKGRVTLKQYPRKKNLFVEVTLTSTGSAVYACGAMNITGKTKYKTLNVHSE 125  
DB 72 GYLSKYSGRANLINFPRNNTFVINIHOITODDGSYKCGIG-TSNQGLSHVSLSEV--- 127  
QY 126 YEPSNEQPMPEPKMFLPYLFQMPAVASSSKFVITVTPAQRGKVP 173  
DB 128 -----SQVPELPSTHIV-----YTKDIGRNVITIECPFKRENVP 160

RESULT 6  
US-08-961-564A-9  
Sequence 9, Application US/08961564A  
Patent No. 6114515  
GENERAL INFORMATION:  
APPLICANT: MU, SHUJIAN  
APPLICANT: SWEET, RAYMOND  
APPLICANT: TRUNER, ALEMEGED  
TITLE OF INVENTION: PIGRL-1, A MEMBER OF IMMUNOGLOBULIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:

```

1 ADDRESSSEE: PATNEP & PRESTIA
2 STREET: P.O. BOX 980
3 CITY: VALLEY PORCH
4 STATE: PA
5 COUNTY: USA
6 ZIP: 19482
7 COMPUTER PEATABLE FORM:
8 MEDIUM TYPE: Diskette
9 OPERATING SYSTEM: DOS
10 SOFTWARE: FastSho for Windows Version 2.0
11 CURRENT APPLICATION DATA:
12 APPLICATION NUMBER: US/08/961,564A
13 FILING DATE: 30-OCT-1997
14 CLASSIFICATION: 536
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER: 60/056,935
17 FILING DATE: 25-AUG-1997
18 ATTORNEY/AGENT INFORMATION:
19 NAME: PRESTIA, PAUL F
20 REGISTRATION NUMBER: 23,031
21 REFERENCE/DOCKET NUMBER: GH-70236
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: 610-407-0700
24 TELEFAX: 610-407-0701
25 TELEX: 846169
26 INFORMATION FOR SEQ ID NO: 9.
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 109 amino acids
29 TYPE: amino acid
30 STRANDEDNESS: single
31 TOPOLOGY: linear
32 MOLECULE TYPE: protein
33 US-08-961-564A-9
34
35 Query Match          9.1% Score 187: DB 3: Length 109:
36 Best Local Similarity 38.3%, Freq No. A Re-10:
37 Matches 41: Conservative 19: Mismatches 41: Indels 6: Gaps 3
38
39 QY      21 PEKAVGEHGGSYTKCPLP----EMWRIYLGRM AGSGTCQTVSTNFIKAEEYKR 75
40       1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
41 Db      4 PE-EVNSVEGNSYSITCYPTPSVNRHKRYW'PdPzAKzSLTLLLSRVSYSKYAR 62
42
43 QY      76 VTIKQVPKKLFLVEFTGLTESNVVVNGCAAMNTTPGKTQKTNV 122
44       1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
45 Db      63 ANLNPHGTGVVNIAQLSDDSCKYKGIGINSIKGLSPDVSLFY 109
46
47 RESULT 7
48 US-C8-434-000A-10
49 : Sequence 10, Application US/0843400UA
50 : Patent No. 6046037
51
52 GENERAL INFORMATION:
53 APPLICANT: ANDREW C. HIATT, JULIAN
54 APPLICANT: K.-C. MA, THOMAS LEINER
55 TITLE OF INVENTION: IMMUNOKILLING CONTAINING PROTECTION
56 TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
57 NUMBER OF SEQUENCES: 19
58 CORRESPONDENCE ADDRESS:
59 ADDRESSEE: Lyon & Lyon
60 STREET: 633 West Fifth Street
61 STREET: Suite 4700
62 CITY: Los Angeles
63 STATE: California
64 COUNTRY: U.S.A.
65 ZIP: 90071
66
67 COMPUTER READABLE FORM:
68 MEDIUM TYPE: 3.5" Diskette, 1.44 MB
69 MEDIUM TYPE: storage
70 COMPUTER: IBM compatible
71 OPERATING SYSTEM: IBM P.C. DOS 5.0
72 SOFTWARE: Word Perfect 5.1
73 CURRENT APPLICATION DATA.
```

```

1 APPLICATION NUMBER: US/08/434,000A
2 FILING DATE:
3 CLASSIFICATION: 435
4 PRIOR APPLICATION DATA: including application
5 PRIOR APPLICATION DATA, described below:
6 APPLICATION NUMBER: 08/367,395
7 FILING DATE: 12/30/94
8 ATTORNEY/AGENT INFORMATION:
9 NAME: Guise, Jeffrey W.
10 REGISTRATION NUMBER: 34,613
11 REFERENCE/DOCKET NUMBER: 212,127
12 TELECOMMUNICATION INFORMATION:
13 TELEPHONE: (619) 552-8400
14 TELEFAX: (619) 552-0159
15 TELEXX: 67-3510
16 TRIMX: SEQUENCE LISTING
17 INFORMATION FOR SEQ ID NO: 10:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 769 amino acids
20 TYPE: amino acid
21 STRANDEDNESS: single
22 TOPOLOGY: linear
23 TOPOLOGY DESCRIPTION: Fat Polyimmuno-globulin Receptor
24 US-08-434-000A-10
25
26 Query Match: 9.0%; Score 185, Dh 3; Length 769;
27 Best Local Similarity 34.5%; Pred. No. 2e-08;
28 Matches 41; Conservative 21; Mismatches 41; Indels 16; Gaps 4;
29
30 GGSATKCPLE---MHVIVLDEMAISGTSIVSTNFIKAERKSGVTKQYPRRN 85
31 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
32 Db 33 GNSVSTICYPDIDSVNRHTRKRYWCROGA-NGYCATLLISSNGLYSKPYSCRASLINPENS 91
33 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
34 Oy 86 LFIIEVETQLESQGYACAGAMNDPRCKTVTLNVHSEVPEPSPWEPQPPEDIPKPHL 144
35 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
36 Ldb 92 TFIIVNHLTQEDTGSYKQSLG-TIRGLTFDVSLEY-----SVEVEPPNDTIV 139
37 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
38
39 RESULT 8
40 US-09-312-157-10
41 Sequence 10, Application US/09312157
42 Patent No. 6303341
43
44 GENERAL INFORMATION:
45 APPLICANT: ANDREW C. HIATT, JULIAN
46 K.-C. MA, THOMAS LEHNKE
47 TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
48 PROTEINS IN PLANTS AND THEIR USES
49 NUMBER OF SEQUENCES: 19
50 CORRESPONDENCE ADDRESS:
51 ADDRESSER: Lyon & Lyon
52 Suite 4700
53 STREET: 633 West Fifth Street
54 CITY: Los Angeles
55 STATE: California
56 COUNTRY: U S A
57 ZIP: 90071
58
59 COMPUTER READABLE FORM:
60 MEDIUM TYPE: 3.5" Diskette, 1.44 MB
61 storage
62
63 COMPUTER: IBM Compatible
64 OPERATING SYSTEM: IBM PC DOS 5.0
65 SOFTWARE: Word Perfect 5.1
66
67 CURRENT APPLICATION DATA:
68 APPLICATION NUMBER: US/09/312,157
69 FILING DATE: 14-May-1999
70 CLASSIFICATION: <Unknown>
71 PRIOR APPLICATION DATA:
72 APPLICATION NUMBER: 08/434,000
73 FILING DATE: <Unknown>
74 ATTORNEY/AGENT INFORMATION:
75 NAME: Guise, Jeffrey W.
76 REGISTRATION NUMBER: 34,613

```



REFERENCE/DOCKET NUMBER: 212/127  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 552-8400  
TELEFAX: (619) 552-0159  
TELEX: 67 351  
SEQUENCE LISTING  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 769 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
DESCRIPTION: Rat Polyimmunoglobulin Receptor  
US-09-312-157-10

Query Match 9 0%: Score 185; DB 4; Length 769;  
Best Local Similarity 34.58; Pred. No. 20-08;  
Matches 41; Conservative 21; Mismatches 41; Indels 16; Gaps 4;

QY 30 GGSVTKCPLE---MIVRIYLCREMASSGTCGVSTNFIKAEYKGVTLKQYPRKN 85  
DB 33 GNSVSTTYPTDTSVNPITPKYWGQA-NIVTATILSSNYLSKESGASLINFPPNS 91  
QY 86 LPIVAVIQLIKSDQVYACAGAMNIDKQTKQVTLNVHSEYEDSWERQPMETPKWPHL 144  
DB 92 TEVINIAHLTQEDTGSYKCGIG-TTNRGLFEDVSLEY-----SQVPEFFNDTHV 139

RESULT 9  
US-09-095-385-4  
Sequence 4, Application US/09095385  
Patent No. 6300104  
GENERAL INFORMATION:  
APPLICANT: Morrison, Sherie L.  
TITLE OF INVENTION: Chittalacharyu, Kore R.  
TITLE OF INVENTION: SECRETORY IMMUNOGLOBULIN PRODUCED  
TITLE OF INVENTION: BY SINGLE CELLS AND METHODS FOR MAKING AND USING  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
STREET: 1150 Santa Monica Boulevard, Suite 400  
CITY: Los Angeles  
STATE: CA  
COUNTRY: USA  
ZIP: 90025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,385  
FILING DATE: 09-JUN-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/050,969  
FILING DATE: 19-JUN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Canady, Karen S.  
REGISTRATION NUMBER: 39,927  
REFERENCE/DOCKET NUMBER: 30435,450SUI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 310 445-1140  
TELEFAX: 310 445-9031  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 608 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-09-095-385-4

Query Match 8 9%: Score 182; DB 4; Length 608;  
Best Local Similarity 37.48; Pred. No. 2-7e-08;  
Matches 43; Conservative 20; Mismatches 40; Indels 12; Gaps 5;

QY 12 PYSCALRIILPHYKVGCGISGVTIKCPPL---FMHVRIYLCREMASSGTCGVSTNFI 67  
DB 21 PLEG-----PE-EVNSVEGNSVSTITPTVSNPHITPKYWGQAAGS-CITLISSEY 73  
QY 68 LKAEYKGVTLKQYPRKNLPIVAVIQLIKSDQVYACAGAMNIDKQTKQVTLNV 122  
DB 74 VSSKYACKANLINFENCIFVNIQAQLSDSDSKYKCGIGINS-KQISHVSLHY 127

RESULT 10  
US-08-434-000A-4  
Sequence 4, Application US/08434000A  
Patent No. 6046037  
GENERAL INFORMATION:  
APPLICANT: ANDREW C. HIATT, TULIAN  
APPLICANT: K.-C. MA, THOMAS LEHNER  
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/434,000A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below: 1  
APPLICATION NUMBER: 08/367,395  
FILING DATE: 12/30/94  
ATTORNEY/AGENT INFORMATION:  
NAME: Guise, Jeffrey W.  
REGISTRATION NUMBER: 34,613  
REFERENCE/DOCKET NUMBER: 212/227  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 552-8400  
TELEFAX: (619) 552-0159  
TELEX: 67-3510  
SEQUENCE LISTING  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 746 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
DESCRIPTION: Human Polyimmunoglobulin Receptor  
US-08-434-000A-4

Query Match 8 9%: Score 182; DB 4; Length 746;  
Best Local Similarity 37.48; Pred. No. 3.6e-08;  
Matches 43; Conservative 20; Mismatches 40; Indels 12; Gaps 5;

QY 12 PVSCLRIILPHYKVGCGISGVTIKCPPL---FMHVRIYLCREMASSGTCGVSTNFI 67  
DB 21 PLEG-----PE-EVNSVEGNSVSTITPTVSNPHITPKYWGQAAGS-CITLISSEY 73

Db 3 PIFG-----PE-EVNSVESNSVSTCYPPPSVNHPRKRW-PQJAPR-CTITLISSEGY 55  
QY 68 IAAEYKGVATLKQYPRKNIPLVEVTLQITPSDSCVYACGAGMNTDQKIQKVTILAV 122  
Db 56 VSSKYGAGRANLINFENGTFFVYVNIADLSQDSGRYKCGGLGINS-RGLSPVSLAV 109

RESULT 11  
US-09-312-157-4  
Sequence 4, Application US/09312157  
Patent No. 6303341

GENERAL INFORMATION:

APPLICANT: ANDREW C. HIATT, JULIAN  
K.-C. MA, THOMAS LEINER

TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
PROTEINS IN PLANTS AND THEIR USES

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" diskette, 1.44 Mb

COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/312,157  
FILING DATE: 14-May-1994

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/434,000  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
NAME: Guise, Jeffrey W.  
REGISTRATION NUMBER: 34,613  
REFERENCE/DOCKET NUMBER: 212/127

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 552-8400  
TELEFAX: (619) 552-0159  
TELEX: 67-351

SEQUENCE LISTING  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 746 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

DESCRIPTION: Human Polymunoglobulin Receptor  
SEQUENCE DESCRIPTION: SEQ ID NO: 4

US-09-312-157-4

Query Match 8.9%: Score 182; DB 4; Length 746;  
Best Local Similarity 37.4%: Pred. No. 3.6e-08;  
Matches 43; Conservative 20; Mismatches 40; Indels 12; Gaps 5;

QY 12 PVSAGALRLPEVKVESELGSGVYIKCPDP-----EMHVRITYLQRMAGSGGTGVYSTTFN 67  
Db 3 PIFG-----PE-EVNSVESNSVSTCYPPPSVNHPRKRW-PQJAPR-CTITLISSEGY 55

QY 68 IAAEYKGVATLKQYPRKNIPLVEVTLQITPSDSCVYACGAGMNTDQKIQKVTILAV 122  
Db 56 VSSKYGAGRANLINFENGTFFVYVNIADLSQDSGRYKCGGLGINS-RGLSPVSLAV 109

RESULT 12  
US-08-434-000A-6  
Sequence 6, Application US/08434000A

Patent No. 6046037  
GENERAL INFORMATION:  
APPLICANT: ANDREW C. HIATT, JULIAN  
K.-C. MA, THOMAS LEINER  
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
PROTEINS IN PLANTS AND THEIR USES  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" diskette, 1.44 Mb

COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/434,000A  
FILING DATE: 12/30/94

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/367,395

ATTORNEY/AGENT INFORMATION:  
NAME: Guise, Jeffrey W.  
REGISTRATION NUMBER: 34,613  
REFERENCE/DOCKET NUMBER: 212/127

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 552-8400  
TELEFAX: (619) 552-0159  
TELEX: 67-3510

SEQUENCE LISTING  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 757 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

DESCRIPTION: Bovine Polymunoglobulin Receptor  
SEQUENCE DESCRIPTION: SEQ ID NO: 6

US-08-434-000A-6

Query Match 8.9%: Score 182; DB 3; Length 757;  
Best Local Similarity 34.1%: Pred. No. 3.7e-08;  
Matches 42; Conservative 21; Mismatches 44; Indels 16; Gaps 4;

QY 17 LRLPEVKVEGEL-----GGSVYIKCPDP-----EMHVRITYLQRMAGSGGTGVYSTTFN 62  
Db 10 LAIFPVYSKMSKSPIFPEEVSVEGRSVSIKYYPPTSVNHRTRKRW-PQJAPR-CTITLISSEGY 68

QY 63 SITNPIKAYKGVATLKQYPRKNIPLVEVTLQITPSDSCVYACGAGMNTDQKIQKVTILAV 122  
Db 69 SSSCYVSDUYVGRANLINFENGTFFVYVNIADLSQDSGRYKCGGLGINS-RGLSPVSLAV 127

QY 123 HSE 125  
Db 128 SQN 130

RESULT 13  
US-09-312-157-6  
Sequence 6, Application US/09312157  
Patent No. 6303341

GENERAL INFORMATION:  
APPLICANT: ANDREW C. HIATT, JULIAN  
K.-C. MA, THOMAS LEINER  
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION

## PROTEINS IN PLANTS AND THEIR USES

NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:

ADDRESS: Lyon & Lyon  
STREET: 633 West Fifth Street  
Suite 4700

CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
Storage

COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/312.157

FILING DATE: 14 May 1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/434,000

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Guise, Jeffrey W

REGISTRATION NUMBER: 34,613

REFERENCE/DOCKET NUMBER: 212/127

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 552-8400

TELEFAX: (619) 552-0159

TELEX: 67-351

SEQUENCE LISTING

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 757 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

DESCRIPTION: Bovine Polymunoglobulin Receptor

US-09-312-157-6

Query Match

Best Local Similarity 34.1%: Score 182, DB 4, Length 757:

Matches 42: Conservative 21, Mismatches 44, Indels 16, Gaps 4,

DB 17 LILPEVAVKPEFL-----CGSVTKCPD-----FMHVRILYLOREMAAGSGTCTGV 62

DB 10 LAIFPVVAKKSPITGPEVSSVEGHSVSKCYPTTSVNRITRYKWRQGA-QSECTITLI 68

DB 63 STTNPIKAFYKQKVTLKQYPRKNIPLVHVTLQTFESDSQVYACGAGMNTDROKTOKVTLNV 122

DB 69 SSGCYVSDYVORANIINPESGTFVVDISHLHKDSQKCKGIGLSS-KCLNIDVSLNV 127

QY 123 ISE 125

DB 128 SOD 130

RESULT 14

US-08-642-406A 22

Sequence 22: Application US/08642406A

Patent No. 5959177

GENERAL INFORMATION:

APPLICANT: Hejin, Mich B.

APPLICANT: Hiatt, Andrew C.

APPLICANT: Ma, Julian K.C.

TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESS: THE SCRIPPS RESEARCH INSTITUTE

STREET: 10666 No 5459177th Torrey Pines Road, Tpc-R

CITY: La Jolla  
STATE: California  
COUNTRY: US

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/642,406A

FILING DATE: 03-MAY-1996

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/591,823

FILING DATE: 02-OCT-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/427,765

FILING DATE: 27-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: Logan, April C.

REGISTRATION NUMBER: 33,950

REFERENCE/DOCKET NUMBER: 184.2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 554-2937

TELEFAX: (619) 554-6312

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 624 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-642-406A-22

Query Match

Best Local Similarity 42.7%: Score 166.5, DB 4, Length 624:

Matches 44: Conservative 11, Mismatches 41, Indels 7, Gaps 4:

DB 30 CGSVTKCPD-----FMHVRILYLOREMAAGSGTCTGVSTNPIKAFYKQVTLKQYPRKNI 85

DB 39 CGSVTKCPD-----FMHVRILYLOREMAAGSGTCTGVSTNPIKAFYKQVTLKQYPRKNI 85

DB 86 LILPEVAVKPEFL-----CGSVTKCPD-----FMHVRILYLOREMAAGSGTCTGV 126

DB 97 LILPEVAVKPEFL-----CGSVTKCPD-----FMHVRILYLOREMAAGSGTCTGV 138

QY 123 ISE 125

DB 128 SOD 130

RESULT 15

US-09-199-534-22

Sequence 22: Application US/09199534

Patent No. 6329569

GENERAL INFORMATION:

APPLICANT: Hejin, Mich B.

APPLICANT: Hiatt, Andrew C.

APPLICANT: Ma, Julian K.C.

TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESS: THE SCRIPPS RESEARCH INSTITUTE

STREET: 10666 No 5459177th Torrey Pines Road, Tpc-R

CITY: La Jolla

STATE: California

COUNTRY: US

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/199,534

FILING DATE: 25-MAY-1998



GenPro version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

UM protein - protein search, using sw model

Run on November 24, 2002 06:00:00, Search time 40 seconds  
(without alignments)  
1299.194 Million cell updates/sec

Title: US-09-651-150b-2  
Perfect score: 2055  
Sequence: 1 MDRLMFLFLPVSALPLI HEPAMMENSSTPTNPVA 390

Scoring table: HLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 13345020 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database

Listing first 45 summaries

1: A\_geneseq\_101002.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted to have a score greater than or equal to the score of the chance being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2047	99.6	390	20	AAV42225
2	2047	99.6	390	20	AAV17496
3	2047	99.6	390	20	AAV05001
4	1157	56.3	432	22	AAE05349
5	732	35.6	255	23	ABD23359
6	550	26.8	107	20	AAV05002
7	410	20.0	97	21	AAV65401
8	198	9.6	771	17	AAW03180
9	195	9.5	771	22	AAV65698
10	185	9.0	769	17	AAW03181

11	185	9.0	769	22	AAV65697	Rat polymeric immu
12	183.5	8.9	532	19	AAV60033	Human immunity rel
13	182	8.9	637	20	AAV34039	Partial amino acid
14	182	8.9	607	20	AAV95601	Human secretory im
15	182	8.9	607	23	AAW47867	Human secretory co
16	182	8.9	746	17	AAW03178	Human poly-immunog
17	182	8.9	757	17	AAW03179	Human poly-immunog
18	182	8.9	774	22	AAV65698	Bovine poly-immunog
19	182	8.9	764	22	AAV65695	Human poly-immunog
20	182	8.9	764	22	AAV65711	Human poly-immunog
21	179	8.7	733	22	AAV65699	Human poly-immunog
22	166.5	8.1	584	12	AAV14670	Truncated poly Ig-
23	166.5	8.1	771	23	AAV65698	Chimeric poly-immu
24	166.5	8.1	774	17	AAW03177	Rabbit poly-immun
25	166.5	8.1	774	22	AAV65700	Rabbit poly-immun
26	161	7.8	307	19	AAW83344	Rat kidney injury
27	156	7.6	584	12	AAV14671	Truncated poly Ig-
28	143.5	7.0	206	20	AAV65696	Human F13b-2 prote
29	142.5	6.9	211	21	AAV65543	Human immune syste
30	131	6.4	746	22	AAV11804	Human immune syste
31	129.5	6.3	332	20	AAV11717	Human protein prote
32	129.5	6.3	332	21	AAV44273	Human protein prote
33	129.5	6.3	332	21	AAV19580	Human protein prote
34	129.5	6.3	332	22	AAV12364	Human protein prote
35	129.5	6.3	332	22	AAV65300	Human protein prote
36	129.5	6.3	332	23	AAV65452	Human protein prote
37	129.5	6.3	332	23	AAV64846	Human protein prote
38	127.5	6.2	203	22	AAV14805	Human bone marrow
39	127.5	6.2	313	22	AAV02639	Human bone marrow
40	127.5	6.2	332	21	AAV67653	Human protein prote
41	122.5	6.0	474	23	AAV14065	Human protein prote
42	122	5.9	298	20	AAV44833	Human protein prote
43	121	5.9	334	19	AAV38332	Human protein prote
44	120.5	5.9	301	20	AAV24032	Human protein prote
45	118.5	5.8	528	15	AAV25952	Human anti-IgM MAB

# ALIGNMENTS

RESULT 1	
AAV42225	
AAV42225 standard; Protein, 390 AA.	
XX	
AC AAV42225;	
XX	
DI 17-DEC-1999 (first entry)	
XX	
DE Human Toso protein sequence.	
XX	
KW Human, Toso, protein, target, drug screening, diagnosis, apoptosis;	
KW apoptosis related disease.	
XX	
OS Homo sapiens.	
XX	
PN W09950671-A2.	
XX	
PD 07-OCT-1999.	
XX	
PE 30-MAR-1999; 99WO-US06945.	
XX	
PR 30-MAR-1998; 98US-0050861.	
XX	
PA (RICE-) RIGEL PHARM INC.	
XX	
PI Payan D.	
XX	
DR WPI: 1999-591379/50.	
XX	
PT N FSDB: AAV25422.	
XX	
PT Screening agents useful for modulating apoptosis and controlling	
XX	
PT apoptosis related diseases	

PS Claim 1; Fig 2a; 75pp; English.  
 XX  
 CC The present invention describes a method of screening for a bioactive  
 CC agent capable of binding a Toso protein. Also described are methods for:  
 CC (1) screening a bioactive agent capable of modulating activity of a  
 CC Toso cell-surface receptor, comprising adding a candidate bioactive  
 CC agent to a cell comprising a recombinant Toso nucleic acid, and  
 CC exposing the cells to an apoptotic agent that will induce apoptosis;  
 CC (2) modulating apoptosis comprising administering an exogenous  
 CC compound that binds Toso, to a cell; (3) identifying a cell containing  
 CC a mutant Toso gene, comprising determining its sequence; (4)  
 CC identifying the Toso genotype, comprising determining the sequence of  
 CC at least one Toso gene; and (5) diagnosing an apoptosis related  
 CC condition, comprising measuring activity of Toso in a tissue, and  
 CC comparing to the activity from non-affected individual's tissue, where  
 CC a reduced activity of the patient indicates risk of an apoptosis related  
 CC condition. The methods are useful for identifying agents capable of  
 CC diagnosing and treating apoptosis related disease, their use for  
 CC modulating apoptosis, and methods for diagnosing the disease state.  
 CC The present sequence represents the human Toso protein for use in  
 CC methods from the invention.  
 CC  
 XX  
 SQ Sequence 390 AA:

Query Match 99.6%; Score 2047; DB 20; Length 390;  
 Best local Similarity 99.7%; Pred. No. 2,1e-166;  
 Matches 389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDRMLPLFLFVSGALRLPEVKEGELGSSVTIKCPLEPMHVRITLCREMGSGTGT 60  
 DB 1 MDRMLPLFLFVSGALRLPEVKEGELGSSVTIKCPLEPMHVRITLCREMGSGTGT 60  
 QY 61 VVSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGYACAGMNTDRGKTUVT 120  
 DB 61 VVSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGYACAGMNTDRGKTUVT 120  
 QY 121 NVHSEYPSMEQPMETPKMFWLHYLFOMPAVYASSSKFVTRVTPAQRCKVPRVHHSSP 180  
 DB 121 NVHSEYPSMEQPMETPKMFWLHYLFOMPAVYASSSKFVTRVTPAQRCKVPRVHHSSP 180  
 QY 181 TQOITHRPVPSVASSVADGKPRTPLPSTTASKISALEGLKQTPSYNHHHTPLHQFALD 240  
 DB 181 TQOITHRPVPSVASSVADGKPRTPLPSTTASKISALEGLKQTPSYNHHHTPLHQFALD 240  
 QY 241 YGSQSGREGOGPHILIPITLGLFLALGLVYKRAVERPRKALSPRARIANVPMALLESSQ 300  
 DB 241 YGSQSGREGOGPHILIPITLGLFLALGLVYKRAVERPRKALSPRARIANVPMALLESSQ 300  
 QY 301 RPRGSPRPSQNNIYSACPRARACADAAGTGAHPYGCAPLPAPLQVSRSPWHAAPSL 360  
 DB 301 RPRGSPRPSQNNIYSACPRARACADAAGTGAHPYGCAPLPAPLQVSRSPWHAAPSL 360  
 QY 361 KTSCEVSLYHQPAAEMEDSDSDYINVA 390  
 DB 361 KTSCEVSLYHQPAAEMEDSDSDYINVA 390

RESULT 2  
 AA17496  
 ID AA17496 standard; Protein; 390 AA.

XX  
 AC AA17496;  
 XX  
 DT 03-AUG-1999 (first entry)  
 XX  
 DE Human Toso protein.

XX Toso protein, tumor necrosis factor mediated apoptosis inhibition;  
 KW TNF mediated apoptosis; T cell overactivity, autoimmune disease;  
 KW Sjorens connective tissue disorder; Transplant rejection; cancer.  
 XX  
 OS Homo sapiens.  
 XX

FH Key Location/Qualifiers  
 FT Peptide 1..17  
 FT Protein 18..390  
 FT Region 254..272  
 FT [label=transmembrane\_region]  
 FN W09925832-A1.  
 PD 27-MAY-1999.  
 XX  
 XX 16-NOV-1998; 98WD-0524391.  
 XX  
 PR 17-AUG-1998; 98US-0135238.  
 PR 17-NOV-1997; 97JN-0066063  
 XX  
 PA (STM) UNIV LEIAND STANFORD JUNIOR.  
 PI Hitoshi Y, Nolan CP;  
 PI  
 DR MPI: 1999-338007/28.  
 DR N-PSDB: AAX76123.  
 XX  
 FT  
 PT  
 PS  
 PS Claim 20; Fig 2a; 70pp; English.

CC The present sequence is a Toso protein (1). (1) has anti-apoptotic  
 CC and cytostatic activity. Toso (named after a Japanese liquor that is  
 CC drunk on New Year's Day to celebrate long life and eternal youth) most  
 CC likely acts by induction of cFLIP expression which inhibits caspase-8  
 CC processing. Recombinant (1) can be used to modulate apoptosis in a cell  
 CC or to treat an apoptosis related condition in a mammal. Apoptosis  
 CC related conditions can also be treated by administration of the Toso  
 CC protein or antibody. Apoptosis related or mediated conditions that can  
 CC be treated include diseases characterized by a cell overactivity, e.g.  
 CC Sjorens connective tissue disorder, autoimmune diseases, diseases where  
 CC T cells actively destroy cells, including transplant rejection and  
 CC conditions where cells of any kind that are not dying express Toso  
 CC appropriately, e.g. cancer of T or B cell origin (where increased  
 CC apoptosis would be appropriate).  
 CC  
 XX  
 SQ Sequence 390 AA:

Query Match 99.6%; Score 2047; DB 20; Length 390;  
 Best local Similarity 99.7%; Pred. No. 2,1e-166;  
 Matches 389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDRMLPLFLFVSGALRLPEVKEGELGSSVTIKCPLEPMHVRITLCREMGSGTGT 60  
 DB 1 MDRMLPLFLFVSGALRLPEVKEGELGSSVTIKCPLEPMHVRITLCREMGSGTGT 60  
 QY 61 VVSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGYACAGMNTDRGKTUVT 120  
 DB 61 VVSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGYACAGMNTDRGKTUVT 120  
 QY 121 NVHSEYPSMEQPMETPKMFWLHYLFOMPAVYASSSKFVTRVTPAQRCKVPRVHHSSP 180  
 DB 121 NVHSEYPSMEQPMETPKMFWLHYLFOMPAVYASSSKFVTRVTPAQRCKVPRVHHSSP 180  
 QY 181 TQOITHRPVPSVASSVADGKPRTPLPSTTASKISALEGLKQTPSYNHHHTPLHQFALD 240  
 DB 181 TQOITHRPVPSVASSVADGKPRTPLPSTTASKISALEGLKQTPSYNHHHTPLHQFALD 240  
 QY 241 YGSQSGREGOGPHILIPITLGLFLALGLVYKRAVERPRKALSPRARIANVPMALLESSQ 300  
 DB 241 YGSQSGREGOGPHILIPITLGLFLALGLVYKRAVERPRKALSPRARIANVPMALLESSQ 300  
 QY 301 RPRGSPRPSQNNIYSACPRARACADAAGTGAHPYGCAPLPAPLQVSRSPWHAAPSL 360  
 DB 301 RPRGSPRPSQNNIYSACPRARACADAAGTGAHPYGCAPLPAPLQVSRSPWHAAPSL 360

QY 361 KISCSEYSLYHOPAAAMEDSDDYINVPA 390  
 DB 361 KISCSEYSLYHOPAAAMEDSDDYINVPA 390

RESULT 3  
 AAY05001  
 ID AAY05001 standard: Protein: 390 AA.  
 XX  
 AC AAY05001:  
 XX  
 D1 16-JUN-1999 (first entry)  
 XX  
 DE Human PIGRL-1 protein sequence.  
 XX  
 KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
 KW X-linked Severe Combined Immunodeficiency; XSCID; Iga deficiency;  
 KW diagnosis; therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP05238-A2.  
 PD 31-MAR-1999.  
 XX  
 PF 14-AUG-1998; 98EP-0306487.  
 XX  
 PR 30-OCT-1997; 97US-0961564.  
 PK 25-AUG-1997; 97US-0056935.  
 XX  
 PA (SMK) SMITHKLINE BEECHAM CORP.  
 PI Sweet RM, Truneh A, Wu S;  
 XX WPI: 1999-192666/17.  
 DR N-PSDB; AAX28178.  
 XX  
 PT New polypeptides encoding PIGRL-1 useful for treating diseases such  
 as X-linked Severe Combined Immunodeficiency  
 PS Claim 11, Page 7; 26pp; English.  
 XX  
 CC This sequence is the human PIGRL-1 protein of the invention.  
 CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
 CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined  
 CC Immunodeficiency (XSCID) and Iga deficiency. These diseases can be  
 CC diagnosed or susceptibility to them predicted by: (1) determining whether  
 CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
 CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
 CC Patients deficient in PIGRL-1 can be treated by administering either the  
 CC PIGRL-1 DNA or its complement of an agonist of PIGRL-1 to the patient  
 CC Patients with excessive expression or activity of PIGRL-1 can be treated  
 CC by administering an antagonist of PIGRL-1, an antisense nucleic acid  
 CC molecule which inhibits the expression of PIGRL-1 or administering  
 CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
 CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
 CC with a candidate compound in the presence of a signal system and noting  
 CC the candidate as an agonist if a signal is produced. The same method can  
 CC be used to identify antagonists of PIGRL-1 but the presence of an  
 CC antagonist is indicated by a decrease in production of the signal.  
 CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
 CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
 CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
 CC This information may then be correlated with the incidence of autoimmune  
 CC disease in those patients to identify whether the mutation causes the  
 CC disease.  
 XX  
 XX Sequence 390 AA:  
 Query Match 99.6%; Score 2047; DB 20; Length 390;  
 Host Local Similarity 99.7%; Prod. No. 2,10-166;  
 Matches 389; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 MDKMLPILYELPYSGALRIIPKVKVGHGELGSGYTIKCPILPMHAKVLYICRHMAGSTGCI 60  
 DB 1 MDKMLPILYELPYSGALRIIPKVKVGHGELGSGYTIKCPILPMHAKVLYICRHMAGSTGCI 60

QY 61 VSTTNF1KAEYKGRVTLKQYPRKNLPLVEVTQITSDSGVYACGAGMTDKQTKVTL 120  
 DB 61 VSTTNF1KAEYKGRVTLKQYPRKNLPLVEVTQITSDSGVYACGAGMTDKQTKVTL 120

QY 121 NVHSEYEPSEOPMPETPKWHLPLYEQMPAYASSKEVTRVITPAQRKQVPHHSSP 180  
 DB 121 NVHSEYEPSEOPMPETPKWHLPLYEQMPAYASSKEVTRVITPAQRKQVPHHSSP 180

QY 181 TQITTHRPVSRASSVAGDKPRITPSTTASKISALGLIKIPOTPSYNNHTIRHQRAD 240  
 DB 181 TQITTHRPVSRASSVAGDKPRITPSTTASKISALGLIKIPOTPSYNNHTIRHQRAD 240

QY 241 YGSGSGRRCOGFHLIPLTIGLFIALLGLVKKRAVERKKALSRHAKRLAVHMKALFSSQ 300  
 DB 241 YGSGSGRRCOGFHLIPLTIGLFIALLGLVKKRAVERKKALSRHAKRLAVHMKALFSSQ 300

QY 301 RPRGSPRPSPSCNNITYSACPPRARGADAAGTGEAPVREGAPLPPAPLVSESPWLHAPSL 360  
 DB 301 RPRGSPRPSPSCNNITYSACPPRARGADAAGTGEAPVREGAPLPPAPLVSESPWLHAPSL 360

QY 361 KISCSEYSLYHOPAAAMEDSDDYINVPA 390  
 DB 361 KISCSEYSLYHOPAAAMEDSDDYINVPA 390

RESULT 4  
 AAE05349  
 ID AAE05349 standard: Protein: 422 AA.  
 XX  
 AC AAE05349:  
 XX  
 DT 12-SEP-2001 (first entry)  
 XX  
 DE Mouse Toso protein.  
 XX  
 KW Mouse; cytotoxic; antineoplastic; immunomodulatory; tissue integrity;  
 KW wound healing; immune response; vaccine; cancer; asthma; allergy;  
 KW cell trafficking; therapy; secreted protein; Fas-induced apoptosis;  
 KW Toso.  
 XX  
 OS Mus sp.  
 XX  
 PN W0200148192-A1.  
 XX  
 PD 05-JUL-2001.  
 XX  
 PF 21-DEC-2000; 2000WO-NZ00256.  
 XX  
 PK 23-DEC-1999; 99US-0171678.  
 PK 28-NOV-2000; 2000US-0724864.  
 XX  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 PI Watson JD, Murison JG;  
 XX  
 DR WPI: 2001-425665/45.  
 DR N-PSDB; AAD10117.  
 XX  
 PT Novel isolated polypeptide useful to isolate corresponding interacting  
 PT proteins or other compounds, to quantitatively determine levels of  
 PT interacting proteins or other compounds, and as therapeutic target  
 PS Claim 6; Page 78-79; 101pp; English.  
 XX  
 CC The patent discloses novel polynucleotides and their corresponding  
 CC proteins which play a major role in induction of growth, cell migration  
 CC and proliferation, cell-cell interaction and the differentiation of  
 CC tissue-specific cells. These proteins are important in the maintenance





diagnosis; therapy.

XX OS Homo sapiens.

XX PN EP905238-A2.

XX PD 31-MAR-1999.

XX PF 14-AUG-1998. 98EP-0306487

XX PR 30-OCT-1997. 97HS-0961564

XX PR 25-AUG-1997. 97HS-0056493

XX PA (SMK ) SMITHKLINE BEECHAM CORP.

XX PI Sweet RW, Truneh A, Wu S;

XX WPI: 1999-192666/17.

XX DR N-PSDB; AAX28179.

XX PT New polypeptides encoding PI3R-1 useful for treating diseases such

XX PT as X-linked severe combined immunodeficiency

XX PS Disclosure, page 9; 26pp; English.

CC This sequence is a fragment of the human PI3R-1 of the invention.

CC Autoimmune diseases involving altered expression or activity of PI3R-1

CC may include Hyper-IGM Immunodeficiency (HIM), X-linked severe combined

CC immunodeficiency (XSCID) and IGA deficiency. These diseases can be

CC diagnosed or susceptibility to them predicted by: (1) determining whether

CC there is a mutation in the genomic copy of the gene encoding PI3R-1; or

CC (2) measuring the amount of PI3R-1 in a sample derived from the patient.

CC Patients deficient in PI3R-1 can be treated by administering either the

CC PI3R-1 DNA or its complement or an agonist of PI3R-1 to the patient.

CC Patients with excessive expression or activity of PI3R-1 can be treated

CC by administering an antagonist of PI3R-1, an antisense nucleic acid

CC molecule which inhibits the expression of PI3R-1 or administering

CC sufficient PI3R-1 to compete with the endogenous activity. PI3R-1 can

CC be used to identify its agonists by contacting a cell expressing PI3R-1

CC with a candidate compound in the presence of a signal system and noting

CC the candidate as an agonist if a signal is produced. The same method can

CC be used to identify antagonists of PI3R-1 but the presence of an

CC antagonist is indicated by a decrease in production of the signal.

CC Antibodies against PI3R-1 may be used to isolate or identify clones

CC expressing PI3R-1. Polynucleotides encoding PI3R-1 may be used to

CC identify chromosomal mutations in the gene encoding PI3R-1 in patients.

CC This information may then be correlated with the incidence of autoimmune

CC disease in those patients to identify whether the mutation causes the

CC disease.

XX SQ Sequence 107 AA:

Query Match 26.8%; Score 550; DB 20; Length 107;

Host local Similarity 99.0%; Pred. No. 2 for 39;

Matches 104; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDRRLMLYLPVSGALRLLPEVKEVEGELGSGVTIKCPLEMHVRYLLCPREMAAGSGTGT 60

DB 1 MDEFLMLYLPVSGALRLLPEVKEVEGELGSGVTIKCPLEMHVRYLLCPREMAAGSGTGT 60

QY 61 VVSTTNFKAEYKGRVTLKQYPRKNLLFVEVDLTESDSGVYACG 105

DB 61 VVSTTNFKAEYKGRVTLKQYPRKNLLFVEVDLTESDSGVYACG 105

RESULT 7

AAW65401

ID AAW65401 standard; Protein: 97 AA.

XX AC AAW65401;

XX DT 01-FEB-2000 (first entry)

Human 5' EST related polypeptide SEQ ID NO:1562.

XX OS Homo sapiens.

XX PN M09953051-A2.

XX PD 21-OCT-1999.

XX PF 09-APR-1999. 99WO-1000712.

XX PR 09-APR-1998. 98US-0057719.

XX PR 28-APR-1998. 98US-0069047.

XX PA (GKST ) GENSET.

XX PI Dumas Mline Edwards J, Dupieret A, Giordano J;

XX WPI: 2000-038446/03.

XX DR N-PSDB; AAZ43015.

XX PT Novel secreted protein 5' expressed sequence tag sequences used in

XX PT diagnostic, forensic, gene therapy, and chromosome mapping procedures

XX PS Claim 3; Page 820-821; 83pp; English.

CC AA42265 to AA43075 represent novel 5' expressed sequence tag (EST)

CC sequences, corresponding to human secreted proteins. AAW64651 to

CC AA46538 represent the EST-related proteins corresponding to AA42265 to

CC AA43052. The 5' ESTs can be used for producing secreted human gene

CC products. They can be used to identify and isolate 5' untranslated

CC regions (UTRs) and upstream regulatory regions which control the

CC location, development stage, rate, and quantity of protein synthesis, as

CC well as stability of mRNA. The ESTs are also useful as probes for

CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can

CC also be used in forensic procedures to identify individuals, or in

CC diagnostic procedures to identify individuals having genetic diseases

CC resulting from abnormal gene expression. The products may also be used in

CC gene therapy protocols. The nucleic acids encoding signal peptides can be

CC used for directing extracellular secretion of a polypeptide or the

CC insertion of a polypeptide into a membrane, or inserting a polypeptide

CC into a cell. The proteins encoded by the EST sequences may be useful in

CC treating a variety of human conditions. Secreted proteins have

CC therapeutic value, and the identification of new secreted proteins is

CC valuable. AA42249 to AA42264 and AAW64644 to AAW64670 represent

CC sequences used in the exemplification of the present invention.

XX SQ Sequence 97 AA:

Query Match 26.8%; Score 550; DB 20; Length 107;

Host local Similarity 99.0%; Pred. No. 2 for 39;

Matches 104; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDRRLMLYLPVSGALRLLPEVKEVEGELGSGVTIKCPLEMHVRYLLCPREMAAGSGTGT 60

DB 1 MDEFLMLYLPVSGALRLLPEVKEVEGELGSGVTIKCPLEMHVRYLLCPREMAAGSGTGT 60

QY 61 VVSTTNFKAEYKGRVTLKQYPRKNLLFVEVDLTESDSGVYACG 105

DB 61 VVSTTNFKAEYKGRVTLKQYPRKNLLFVEVDLTESDSGVYACG 105

RESULT 8

AAW03180

ID AAW03180 standard; Protein: 771 AA.

XX AC AAW03180;

XX DT 24-FEB-1997 (first entry)

```

XX Mouse poly-immunoglobulin receptor.
DE
XX
XX Mouse, immunoglobulin, receptor, protection protein, mutants;
KM heavy chain; antigen binding domain, protection; pathogen;
KM mucosal, environment; gastrointestinal; passive; immunisation;
KM Guy's 13 antibody; prevention; dental caries; Streptococcus;
KM poly. sorbinus; murine.
XX
OS Mus musculus.
XX
XX Key Location/Qualifiers
XX Region 13..45
XX /note= "putative immunoglobulin binding residues
XX of domain I"
XX
XX Domain 1..120
XX /label= domain_I
XX
XX Domain 110..230
XX /label= domain_II
XX
XX Domain 210 340
XX /label= domain_III
XX
XX Domain 320..450
XX /label= domain_IV
XX
XX Domain 440..550
XX /label= domain_V
XX
XX Region 550..606
XX /note= "external portions of domain VI"
XX
XX Region 550..627
XX /note= "external portions of domain VI"
XX
XX Region 625..660
XX /label= transmembrane-segment
XX
XX Region 650..771
XX /label= intracellular-portion
XX
XX WO9621012-A1.
XX
XX 11-JUL-1996.
XX
XX 27-DEC-1995; 95WO-US16889.
XX
XX 04-MAY-1995; 95US-0434000
XX 30-DEC-1994; 94US-0367395.
XX
XX (PLAN-) PLANT BIOTECHNOLOGY INC.
XX (UNM-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
XX (PLAN-) PLANET BIOTECHNOLOGY INC.
XX
XX Hiatt AC, Lehner T, Ma JKC;
XX
XX WPI: 1996-333987/33.
XX N-PSDB; AAT31290.
XX
XX Immunoglobulin and protection protein complex and its produ. in
XX plants - useful for passive immunisation against mucosal antigens,
XX esp. against S. mutans and S. sorbinus to prevent dental caries
XX
XX
XX Disclosure: pages 117 121, 153pp; English.
XX
XX The present sequence is the mouse poly-immunoglobulin (Ig)
XX receptor, a portion of which corresp. to residues 1-627, pref.
XX 1-606, or esp. residues 13-45, 110-230, 210-340, 320-450,
XX 440-550, 550-606 or 550-627 comprises a protection protein (PP).
XX The Ig of the invention comprises a PP as above in association with
XX an Ig derived heavy chain, having at least a portion of an antigen
XX (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
XX gastrointestinal, environments, therefore enhancing its
XX effectiveness in passively immunising animals against mucosal
XX pathogens. The Ag binding domain is specifically derived from the
XX Guy's 13 antibody, and the Ig can be used to prevent dental caries
XX by binding, e.g. Streptococcus mutans serotypes C, E and F, or
XX S. sorbinus serotypes d and g.
XX
XX Sequence 771 AA:

```

```

Query Match 9.64; Score 198; DB 17; Length 771;
Best local Similarity 31.58; Pred. No. 4,1e-08;
Matches 53; Conservative 23; Mismatches 68; Indels 24; Gaps 5;
UY 10 FLVPSALRIPPEKVEELGASVTKRPLPE-----MHVPTVTPKMAHSGTGTAVST 65
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13 FSGVSKSPILGPGQEVSEIDSVSLTCVYPDTVNHRTKYMFGSA-SGMETLISNN 71
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 66 NPIKAPYGRVTLKQYPRKNLFLVEVYQILHESGVYACGAGANNIDRGTQKTKLNVHSE 125
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 72 GLSKSEYSGKANLINFENNTFVNIQLTVDGTSKQDLG-TSMRGSEFVSLV--- 127
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
UY 126 YEPSMEQPMPEITPKMFLPYLFQMPAYASSSKFVTAVTTPAQRKVP 173
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 128 -----SOVPLPSDTHV-----YTDKIDGNVTHPEPRKRNVP 160
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
AAG5698
ID AAG5698 standard; protein; 771 AA.
XX
XX AAC65698;
XX
XX 07-JAN-2002 (first entry)
XX
XX Mouse polymeric immunoglobulin receptor (pIgR) sequence.
XX
XX polymeric immunoglobulin receptor, pIgR, ligand; therapeutic;
XX carcinoma diagnosis; veterinary; mouse.
XX
XX Mus sp.
XX
XX WO200172846-A2.
XX
XX 04-OCT-2001.
XX
XX 26-MAR-2001; 2001WO-US09699.
XX
XX 27-MAR-2000; 2000US-192197P.
XX 27-MAR-2000; 2000US-192198P.
XX
XX (RESC ) UNIV CALIFORNIA.
XX
XX Mostov KK, Chapin SJ, Richman-Wisenstat J;
XX
XX WPI: 2001-611619/70.
XX
XX New ligands binding to a specific region of a polymeric immunoglobulin
XX receptor, useful for transporting therapeutic or diagnostic
XX compositions into or across cells expressing pIgR e.g. in drug delivery
XX
XX
XX Disclosure: Fig 1; 102pp; English.
XX
XX The invention provides ligands that bind specifically to a region of an
XX animal cell polymeric immunoglobulin receptor (pIgR). The pIgR cleaves
XX to produce a stalk region remaining attached to the cell and a secretory
XX component existing in the organ of interest in several forms. The ligands
XX do not bind to the stalk or the most abundant form of the secretory
XX component present in the organ under physiological conditions. The
XX ligands are useful for transporting therapeutic or diagnostic
XX compositions into or across cells expressing pIgR, useful to introduce
XX or transport ligands such as antibodies and/or to deliver biologically
XX active components such as proteins, nucleic acids or detectable labels.
XX They are used to deliver therapeutic compositions to mucosal surfaces
XX such as the gastro-intestinal tract, respiratory system etc. in humans.
XX They are also useful to label cells expressing pIgR, e.g. to distinguish
XX epithelial cells from a mixed cell population in pathology studies or to
XX aid in carcinoma diagnosis (since pIgR expression is reduced in
XX carcinomas relative to normal epithelium). They can also be used to
XX deliver veterinary compositions, especially in mammals such as farm,
XX domestic or wild mammals or birds e.g. birds reared for human
XX
XX

```

[illegible]

receptor, useful for transporting therapeutic or diagnostic compositions into or across cells expressing pigK e.g. in drug delivery

Disclosure: Fig 1: 102pp; English.

The invention provides ligands that bind specifically to a region of an animal cell polymeric immunoglobulin receptor (pIgR). The pIgR cleaves to produce a stalk region remaining attached to the cell and a secretory component existing in the organ of interest in several forms. The ligands do not bind to the stalk or the most abundant form of the secretory component present in the organ under physiological conditions. The ligands are useful for transporting therapeutic or diagnostic compositions into or across cells expressing pIgR, useful to introduce or transport ligands such as antibodies and/or to deliver biologically active components such as proteins, nucleic acids or detectable labels. They are used to deliver therapeutic compositions to mucosal surfaces, such as the gastro-intestinal tract, respiratory system etc. in humans. They are also useful to label cells expressing pIgR, e.g. to distinguish epithelial cells from a mixed cell population in pathology studies or to aid in carcinoma diagnosis (since pIgR expression is reduced in carcinomas relative to normal epithelium). They can also be used to deliver veterinary compositions, especially to mammals such as farm, domestic or wild mammals or birds e.g. birds reared for human consumption. The present sequence represents a rat pigK sequence.

Sequence 769 AA:

Query Match 9.0%, Score 185, DB 22, Length 769;  
Best Local Similarity 34.5%, Prod. No. 5.2e-07;

Matches 41; Conservative 21; Mismatches 41; Indels 16; Gaps 4.

30 GGSVITKCPLE---MIVRTYLCREMAAGSTGIVSTNFIKAEYKGVTLKQYRKRN 85  
DB 33 GNSVSTCYTPPTSVNRTFRYWCQGA-NGYCATLISNGLSKSEYSGRASLTNPENS 91  
QY 86 LPLVEVTOITTESDGVYACGAGMNTDRCKTQKVTILNVHSEVEPSPWEPMPETPKMPL 144  
DB 92 TVVINAHITDIEDTSGYKCGLS-TTNRGLFPDVSLEV-----SQVPEPNDTHV 139

RESULT 12

AAW50033  
ID AAW50033 standard; Protein, 532 AA.

XX AAW50033;

DT 26-JUN-1998 (first entry)

XX Human immunity related factor.

XX Lymph node; human; immunity related factor; research; treatment;

XX Immune disease; infectious disease.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..16

FT /label- sig-peptide 17..532

FT Peptide /label- mat-peptide

PN JPI0072495-A.

PD 17-MAR-1998.

XX 11-JUN-1997; 97JP-0153218.

XX 13-JUN-1996; 96JP-0152362

XX (ASAH ) ASAH KASEI KOGYO KK.

XX WPI: 1998-234766/21.

N-PSUB; AAV20383.  
XX Immunity related factor - useful in the treatment of immune related  
PT and infectious diseases

XX Claims 1, 3 and 4; Pages 18-20; 21pp; Japanese.

XX The present sequence is a lymph node derived human immunity  
CC related factor, which can be used to research and treat immune and  
XX infectious diseases.

Sequence 532 AA:

Query Match 8.9%, Score 193.5; DB 14; Length 532;  
Best Local Similarity 44.0%, Prod. No. 4.4e-07;

Matches 93; Conservative 56; Mismatches 147; Indels 109; Gaps 18;

3 KWLWPLLYPLVSGALRIKILPEVK-----VEGHLGGSVITKCP 38  
DB 25 KWLWE-GSIPKRTILKAMGTLRPSSPLCWRESSPAAPNSLKSGRIYSGPGCAVTLQCH 83

QY 39 LP-----EMIVRTYLCREMAAGSTGIVSTNFIKAEYKGVTLKQYRKRNFLVEVTOU 94

DB 84 YAPSSVNRHQRKYRCRIQPPWITCOTIVSTNQTTHHXYKQRYVALIDPFGQLFVVKLSOL 143

QY 95 TESDGVYACGAGMNTDRCKTQKVTILNVHSEVEPSPWEPMPETPKMPL 153

DB 144 SPDDIGCYLCIGI---SENNMLFLSMNLISAGPA---STLPATAPAGEL---TMRSY 193

QY 154 ASSSKFVIRKVTLPACQKAVPPVHHSSPTQITTHRPVYSKASVAGCK--KRTPLPSTTAS 211

DB 194 GIASVYANRWIP---FTTQTLISQGTAMVIVASVTPGSKTASAPGPRPGATRPAPGT 249

QY 212 KISALEYTLK-----PRTFSYNIHTRLHPQPALDYNSQSGSEYGFHLLPTLDFLLA 266

DB 250 G-SMAEGSVKAPAPIPSPSPSKSMSNTIEGVWEGTRSS----- 288

QY 267 LALJLVYKKA---VEKKAKSKRKK-----LAVMKALSSQK----- 301

DB 289 ---VTNNARASKDRRETTTKADRPREDIEGVRI-ALDAKAKVLTGIGPALVSETLAW 343

QY 302 ---PRGSP--RPRSONNIYSACPRRAGCAADAGCGAPVPGCAP 341

DB 344 KILPQATIVYSKQSDSGSICETTP--AAGMWILGTPAADVWILGTP 386

RESULT 13

AAV34099  
ID AAV34099 standard; Protein, 607 AA.

XX AAV34099;

DT 20-DEC-1999 (first entry)

XX Partial amino acid sequence of plasmid pSHUSC.

XX Multimeric protein; immunoglobulin; receptor; ligand complex;

XX hetero-dimeric receptor; transmembrane protein; transgenic.

XX Synthetic.

PN W09949024-A2.

PD 30-SEP-1999.

XX 24-MAR-1999; 99WO-US06506.

XX 25-MAR-1998; 98US-0079249.

XX (PLAN-) PLANET BIOTECHNOLOGY INC.

XX Wycoff Kt., Jaiswal SK;

XX

XX

XX

XX

DR WPI: 1999-580446/49.  
 DR N-PSDB: AAM72220.  
 XX  
 PT Producing heterologous multimeric proteins in plants, transformed with  
 PT several plasmids expressing polypeptide components, particularly for  
 PT immunoglobulins  
 XX  
 PS Example 1: Fig 8: 42pp; English.  
 CC The invention relates to a method for producing heterologous, multimeric  
 CC proteins in plant cells. The method comprises: (a) transforming the cells  
 CC with several naked plasmids each encoding some, but not all, of the  
 CC polypeptide components of the multimeric proteins, and together providing  
 CC all the polypeptide components; and (b) culturing the cells. The method  
 CC is used to produce biologically active multimeric proteins particularly  
 CC immunoglobulins, receptor ligand complexes, homo- or heterodimeric  
 CC receptors, or trimeric proteins. This method provides properly  
 CC associated and assembled multimeric proteins in a fast and efficient  
 CC process, without the need to cross plants expressing single component of  
 CC the protein. Transgenic plants containing adjacent and stably integrated  
 CC plasmids, and their progeny can also express the multimeric proteins. The  
 CC present sequence represents the partial amino acid sequence of the  
 CC plasmid pSHSC.  
 XX  
 SQ Sequence 607 AA:  
 Query Match 8.9%; Score 182; DB 20; Length 607;  
 Best Local Similarity 37.4%; Pred. No. 6.9e-07;  
 Matches 43; Conservative 20; Mismatches 40; Indels 12, Gaps 5;  
 QY 12 PVSGALRIIPKVEKVEGELGSSVTIKPTFF---EMHVRILYCFEMAGSGTCTGVSTTNF 67  
 DB 21 PIFG-----PE-EVNSVEGNSVSIITCYPPISVNRHTKRYMCRQCAROG-CITLISSEY 73  
 QY 68 IKAHYGPRVTLKQYPRKNIPLVEVPTQLTPSDSGVYACGAGMNTDCKTQKTLNV 122  
 DB 74 VSKRYAGRANLTNPNNGTFVYVNIAGLSQDSGRYKCGIGINS-RGLSPDVSLEY 127  
 DE 74 VSKRYAGRANLTNPNNGTFVYVNIAGLSQDSGRYKCGIGINS-RGLSPDVSLEY 127  
 RESULT 14  
 AAM95601  
 ID AAM95601 standard; Protein: 607 AA.  
 AC AAM95601;  
 XX  
 DT 08-JUN-1999 (first entry)  
 XX  
 DE Human secretory Immunoglobulin A component.  
 XX  
 KW Immunoglobulin A; secretory; component; IgA; human; treatment;  
 KW prevention; infection; HIV; AIDS; cold; flu; virus;  
 KW human immunodeficiency virus; respiratory syncytial virus.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09657993-A1.  
 PD 23-DEC-1998.  
 XX  
 PF 10-JUN-1998: 98WO-0511975  
 XX  
 PR 19-JUN-1997: 97NS-0050969  
 XX  
 PA (RBC) UNIV CALIFORNIA.  
 XX  
 PI Chintalacheruvu KR, Morrison SL;  
 XX  
 DR WPI: 1999-080950/07.  
 DR N-PSDB: AAK07407.  
 XX  
 PT Producing secretory immunoglobulin in single cells - useful to  
 PT produce commercial quantities of secretory immunoglobulin to prevent  
 PT of treat infections

XX  
 PS Disclosure, Pages 22-24, 39pp, English.  
 CC The sequence is that of the secretory component of human secretory  
 CC immunoglobulin A (SIgA). It can be used as part of a method for  
 CC the production of SIgA molecules. This method is useful for  
 CC producing commercial quantities of SIgA (especially SIgA) to treat  
 CC or prevent infections, in particular SIgA produced by the method  
 CC can be used to prevent or treat infections in mammals, birds or  
 CC fish, especially systemic infections or infections at a mucosal  
 CC surface. It is especially useful to prevent or treat infection  
 CC with human immunodeficiency virus (HIV), respiratory syncytial  
 CC virus, flu virus or cold virus. The method allows production of  
 CC commercial quantities of SIgA molecules for therapeutic use, not  
 CC previously possible; production using non-plant cells and a  
 CC single cell type is more efficient than a previous multi-step  
 CC process of fusing recombinant plant cells, and avoids alterations  
 CC of the SIgA by plant cells. SIgA molecules are more stable  
 CC and resistant to proteolysis than previously used IgA molecules,  
 CC and can be administered to prevent as well as to treat infections,  
 CC unlike e.g. IgG and IgM molecules.  
 XX  
 SQ Sequence 607 AA:  
 Query Match 8.9%; Score 182; DB 20; Length 607;  
 Best Local Similarity 37.4%; Pred. No. 6.9e-07;  
 Matches 43; Conservative 20; Mismatches 40; Indels 12, Gaps 5;  
 QY 12 PVSGALRIIPKVEKVEGELGSSVTIKCPD---EMHVRILYCFEMAGSGTCTGVSTTNF 67  
 DB 21 PIFG-----PE-EVNSVEGNSVSIITCYPPISVNRHTKRYMCRQCAROG-CITLISSEY 73  
 QY 68 IKAHYGPRVTLKQYPRKNIPLVEVPTQLTPSDSGVYACGAGMNTDCKTQKTLNV 122  
 DB 74 VSKRYAGRANLTNPNNGTFVYVNIAGLSQDSGRYKCGIGINS-RGLSPDVSLEY 127  
 DE 74 VSKRYAGRANLTNPNNGTFVYVNIAGLSQDSGRYKCGIGINS-RGLSPDVSLEY 127  
 RESULT 15  
 AAM47867  
 ID AAM47867 standard; Protein: 607 AA.  
 AC AAM47867;  
 XX  
 DT 22-FEB-2002 (first entry)  
 XX  
 DE Human secretory component.  
 XX  
 KW Human; immunoadhesin; intercellular adhesion molecule; ICAM-1;  
 KW human rhinovirus; immunoglobulin heavy chain; J chain; IFV; common cold;  
 KW transgenic plant.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W02090183529-A2.  
 PD 08-NOV-2001.  
 XX  
 PF 28-APR-2001: 2001WO 0513932.  
 XX  
 PR 28-APR-2000: 2000US 209298P.  
 XX  
 PA (PLAN-) PLANET BIOTECHNOLOGY INC.  
 XX  
 PI Patrick JW, Wycoff KL;  
 XX  
 DR WPI: 2002-04181/05.  
 DR N-PSDB: ABA05260.  
 XX  
 PT Immunoadhesin for treating human rhinovirus infection comprises  
 PT chimeric intercellular adhesion molecule-1, and optionally a J chain  
 PT and secretory component in association  
 XX  
 PS Examples: Fig 8: 138pp; English.











diagnosis: therapy.

XX Homo sapiens.

XX kb905238-A2.

XX 31-MAR-1999.

XX 14-AUG-1998; 98EP-0306487.

XX 30-OCT-1997; 97US-0961564.

XX 25-AUG-1997; 97US-0056444.

XX (SMK ) SMITHKLINE BEECHAM CORP.

XX Sweet RW, Trunch A, Wu S:

XX WPL: 1999-192556/17.

XX N-PSDB: AMX28178.

XX New polypeptides encoding PIGRL-1 useful for treating diseases such as X-linked Severe Combined Immunodeficiency

XX Claim 11: Page 7; 26pp: English.

This sequence is the human PIGRL-1 protein of the invention.

CC Autoimmune diseases involving altered expression or activity of PIGRL-1 may include Hyper-IgM immunodeficiency (HIM), X linked Severe Combined Immunodeficiency (XSCID) and IgA deficiency. These diseases can be diagnosed or susceptibility to them predicted by: (1) determining whether there is a mutation in the genomic copy of the gene encoding PIGRL-1; or (2) measuring the amount of PIGRL-1 in a sample derived from the patient. CC Patients deficient in PIGRL-1 can be treated by administering either the PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient. CC Patients with excessive expression or activity of PIGRL-1 can be treated by administering an antagonist of PIGRL-1, an antisense nucleic acid molecule which inhibits the expression of PIGRL-1 or administering sufficient PIGRL-1 to compete with the endogenous actively PIGRL-1 CC CC be used to identify its agonists by contacting a cell expressing PIGRL-1 with a candidate compound in the presence of a signal system and noting CC the candidate as an agonist if a signal is produced. The same method can be used to identify antagonists of PIGRL-1 but the presence of an CC antagonist is indicated by a decrease in production of the signal. CC Antibodies against PIGRL-1 may be used to isolate or identify clones CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients. CC This information may then be correlated with the incidence of autoimmune CC disease in those patients to identify whether the mutation causes the CC disease.

XX Sequence 390 AA:

Alignment Scores:

Pred. No.: 1,4e-165 Length: 390  
Score: 2056.00 Matches: 390  
Percent Similarity: 100.00% Conservative: 0  
Host local Similarity: 100.00% Mismatches: 0  
Query Match: 58.91% Indels: 0  
Dbs: 20 Gaps: 0

08-09-651-150b-1 (1-1911) x AA05001 (1-390)

QY 74 ATGGACTTGTGGCTTGGCCACTTACTCTGCGCAGTACAGGGCCGTCAGATCCTC 133  
DB 1 MetAspSerHisPheLeuThrProLeuThrPheLeuProValSerCylAlaLeuArgLleLeu 20  
QY 134 CTAAATTAAGSTAAAG 193  
DB 21 ProGluValIleValCylCylCylCylCylCylCylCylCylCylCylCylCylCylCyl 40  
QY 194 GAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 253  
DB 41 GluMetHisValArgLleTyrLeuGlySerArgGluMetAlaGlySerGlyThrCysGlyThr 60

QY 254 GTGATATGCAACCAACATTCATCAAGAGTACAGATATCAAGAGAGAGAGAGAGAGAGAGAG 313  
DB 61 ValValSerThrThrAsnPhelIleGlySalActuThrGlyGlyArgValThrLeuGlyGln 80  
QY 314 TACCCAGCCAGCAAGATTTCTTTTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 373  
DB 81 TyrProArgGlyAsnLeuPheLeuValGluValThrGlnLeuThrGlnSerAspSerGly 100  
QY 374 GCTATGCTTGGTGGAG 433  
DB 181 ValTyrAlaCysGlyAlaGlyIleMetAsnThrAspArgGlySerGlyThrGlyValThrLeu 120  
QY 434 AATGCCACAGGATACAG 493  
DB 121 AsnValHisSerCylThrCylProSerThrProCylCylCylCylCylCylCylCylCyl 140  
QY 444 TAAATTCATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 553  
DB 141 TrpPheHisLeuProTyrLeuPheLeuMetProAlaTyrAlaSerSerSerLysPheVal 160  
QY 554 ACTATAGTTACAGATACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 613  
DB 161 ThrArgValThrThrProAlaGlnArgGlyValProProValHisHisSerSerPro 180  
QY 614 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 673  
DB 181 ThrThrGlnIleThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLys 200  
QY 674 GCGCCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 733  
DB 201 ProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGlnGlyLeuLeu 220  
QY 734 AAGCCCGAG 793  
DB 221 LysProGlnThrProSerTyrAsnHisHisThrArgLeuHisArgGlnAlaLeuAsp 240  
QY 794 TATGATCACTCTGAG 853  
DB 241 TyrCysGlnSerCylArgCylCylCylCylCylCylCylCylCylCylCylCylCylCyl 260  
QY 854 AAGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 913  
DB 261 GlyLeuPheLeuAlaLeuLeuGlyLeuValValLysValAlaValLysValAlaValLys 280  
QY 914 GATTCCTTAAAG 973  
DB 281 AlaLeuSerAlaGlyAlaArgAlaArgLeuAlaValAlaGlnAlaLeuHisSerSerGln 300  
QY 974 AGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1033  
DB 301 ArgProArgGlySerProArgProArgSerCylHisAsnIleGlySerAlaCysProArg 320  
QY 1034 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1093  
DB 321 ArgAlaAlaGlyAlaAlaSerAlaAlaGlyThrGlyAlaLeuProValProGlyProGlyAla 340  
QY 1094 CCGTTCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 1153  
DB 341 ProLeuProProAlaProLeuGlnValSerCylHisSerCylProThrPheHisAlaProSerLeu 360  
QY 1154 AAGACCACTTGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1213  
DB 361 LysThrSerCysGluTyrValSerLeuGlyHisGlnProAlaAlaMetMetCylAspSer 380  
QY 1214 GATTCAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1273  
DB 381 AspSerAspAspTyrIleAsnValProAla 390

RESULT 4  
AA05349  
ID AA05349 standard: Protein: 422 AA.



















01	07-JAN-2002	(first entry)	
CE	Human polymeric immunoglobulin receptor (PIgR) sequence.		
CE	Human polymeric immunoglobulin receptor (PIgR); ligand, therapeutic,		
KW	carcinoma diagnosis; veterinary; human.		
KW			
XX	Homo sapiens.		
XX			
FT	Key	Location/Qualifiers	
FT	Peptide	487..603	
FT	/note="peptide to which a ligand binds to (claim 8)"		
FT	Peptide	487..607	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	487..611	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	487..615	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	487..618	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	520..607	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	520..611	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	520..615	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	520..618	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	544..615	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	544..618	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	560..607	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	560..611	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	560..615	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	560..618	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	574..607	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	574..611	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	574..615	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	574..618	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	577..607	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	577..611	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	577..615	
FT	/note="peptide to which a ligand binds to"		
FT	Peptide	577..618	
FT	/note="peptide to which a ligand binds to"		
XX	WO200172846-A2.		
XX	04-OCT-2001.		
XX	26-MAY-2001; 2001WO-0509699.		
XX	27-MAY-2000; 2000US 192197P.		
XX	27-MAY-2000; 2000US-192198P		
XX	(REGC ) UNIV CALIFORNIA.		
XX			
XX	Mostov KE, Chaplin SJ, Richman-Bissonstat J;		

[illegible]









Qy 1343 GAGCCCTTCATCAGTTCCCATGCCCATCTCGACT 1376  
 : ||| ||||| ||| :  
 Db 461 sValProCysHisPheProCysLysPheSerSer 472

Search completed: November 24, 2002, 05:54:18  
 Job time : 79.5 secs



TELEPHONE: 610-407-0700

TELEFAX: 610 437 0701  
TELEX: 846165  
INFORMATION FOR SEQ ID NO: 2  
SEQUENCE CHARACTERISTICS:  
LENGTH: 390 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-564A-2

Alignment Scores:

Pred. No.:	1,71e+168
Score:	2056.00
Percent Similarity:	100.00%
Best local Similarity:	100.00%
Query Match:	58 91%
DB:	3
	Gaps: 0
	Matches: 390
	Conservative: 0
	Mismatches: 0
	Indels: 0

US-09-651-150B-1 (1-1911) x US-08-961-564A-2 (1-390)

[illegible][illegible]

## RESULTS

US-09-724-864-45  
; Sequence 45, Application US/09724864  
Patent No. 630253

FILE# NO. 058036Z  
; GENERAL INFORMATION  
; APPLICANT: WATSON

APPLICANT: Murison, James G.  
TITLE OF INVENTION: Polynucleotides, polypeptides expressed

```

; TITLE OF INVENTION:  by the polynucleotides and methods
; FILE REFERENCE:  11000.105001

```

```

; CURRENT A
; CURRENT H
; CURRENT A

```

NUMBER OF SEQ TR

; SOFTWARE: Fa  
; SEQ ID NO 45

```

; LENGTH: 422
;
; TYPE: PRT
;

```

```

US-09-651-150b-1 (1-1911) x US-09-724-864-45 (1-4222)
US-09-724-864-45
Organism: Mouse
Alignment Scores:
  Freq. No.: 1,75e+94
  Score: 1194.50
  Percent Similarity: 68.32%
  Best Local Similarity: 57.45%
  Query Match: 34.23%
  DB: 4
  Gaps:
  Length:
  Matches:
  Mismatches:
  Indels:

```

QY	74	ATGACTTCCTGGATTTGGACATTTATTTCTCTGCAATATATAGAGAGAGCTTCTC	133
Db	1	McEAspPhePheTyrPheLeuLeuTyrPheLeuLeuValSerGlyAlaLeuArgValLeu	20
QY	134	CGAGAGCTAAAGCTAGACAGCGGAGCTGGGCGAGACATTTACATGCAAAAGCCATCTTCT	193
Db	21	ProGluValGlnLeuAsnValGluTyrPheValSerLeuLeuLeuLysProLeuPro	40
QY	194	GAATGCGATCTGAGAGATATCTGTGCGCGGAGACAGGCTGGGAGCTGGGAACCATGTGTACG	253
Db	41	GlnLeuGlnValAlaGlnMetTyrLeuGlnGlyArgGlnIleMcEAlaLysProGlyLysGlySerThr	60
QY	254	GTGATTTCTTCACATCAATTTATAGAGCAAAATTAAGAGAGAGCAATTTCTTGAGAGAA	313























```

01 Y 1408 GCGTCGCCCCCAGCGTGGCT-----GGACACCTCCGC 1441
02 ||| ||||| |||
03 Db 444 nProtoprotyrGlyAlaqlLeuAlaAsnSerAsnAlaHisProGly 460
04
05 RESULT 15
06 US-08-642-406A-22
07 : Sequence 22, Application: US/0844406A
08 : Patent No. 5959177
09
10 GENERAL INFORMATION:
11 APPLICANT: HeIn, Mich B.
12 APPLICANT: Hlatc, Andrew C.
13 APPLICANT: Ma, Julian K.C
14 TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSHMLED
15 TITLE OF INVENTION: SECRETORY ANTIBODIES
16 NUMBER OF SEQUENCES: 26
17
18 CORRESPONDENCE ADDRESS:
19 ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
20 STREET: 10666 No. 5959177th Torrey Pines Road, TPC-8
21 CITY: La Jolla
22 STATE: California
23 COUNTRY: US
24 ZIP: 92037
25
26 COMPUTER READABLE FORM:
27 MEDIUM TYPE: floppy disk
28 COMPUTER: IBM PC compatible
29 OPERATING SYSTEM: PC-DOS/MS-DOS
30 SOFTWARE: PatentIn Release #1.0, Version #1.25
31
32 CURRENT APPLICATION DATA:
33 APPLICATION NUMBER: US/08/642,406A
34 FILING DATE: 03-MAY-1996
35 CLASSIFICATION: 800
36
37 PRIOR APPLICATION DATA:
38 APPLICATION NUMBER: US 07/591,823
39 FILING DATE: 02-OCT-1990
40
41 PRIOR APPLICATION DATA:
42 APPLICATION NUMBER: US 07/427,765
43 FILING DATE: 27-OCT-1989
44 ATTORNEY/AGENT INFORMATION:
45 NAME: Logan, April C.
46 REGISTRATION NUMBER: 33,950
47 REFERENCE/DOCKET NUMBER: 184_2
48 TELECOMMUNICATION INFORMATION:
49 TELEPHONE: (619) 554-2937
50 TELEFAX: (619) 554-6312
51 INFORMATION FOR SEQ ID NO: 22:
52 SEQUENCE CHARACTERISTICS:
53 LENGTH: 624 amino acids
54 TYPE: amino acid
55 TOPOLOGY: linear
56 MOLECULE TYPE: protein
57 US-08-642-406A-22
58
59 Alignment Scores:
60 Pred. No.: 3.34e-06 Length: 624
61 Score: 167.50 Matches: 52
62 Percent Similarity: 53.85% Conserved: 18
63 Best Local Similarity: 40.00% Mismatches: 46
64 Query Match: 4.80% Indels: 14
65
66 DH: 2 Gaps: 6
67
68 US-09-651-150H-1 (1-1911) x US-08-642-406A-22 (1-624)
69
70 QY 85 GTTTGGACTATTATTT---CTGCAAGTAAGAAAGCCCTGAGAGATCTCCAGAGT 141
71 ||| :::::||||| |||:::||||| ||||| ::::::
72 Db 18 AladInSerSerIleuclgylProSerSeriIlePhcglyProGlyValaAsnValle 37
73
74 OY 142 AAAGGTACAGAGCGAGCTGGGCGCATGTTACCATCAATATGCCCACTTCCN----- 193
75 ::::: ||| |||||:::||| |||
76 Db 37 uGIU-----GLYASpserValserIleIlehcysIyTYProIlnurse 52
77
78 Y 194 -----GAATCATGTCAGCATATAATATTCCTGCGCGAGATGCTGATCTGCAACATGTGC 249
79 ||| ||| :::: ||||||||| ||||| |||

```



Sequence 4, Appl  
Sequence 59, Appl  
Sequence 124, App  
Sequence 1, Appl

CLASSIFICATION, and  
PRIOR APPLICATION DATA

5	200	5.7	771	9	US-09-982-107-8	Sequence 8, Appl1
6	199	5.7	746	9	US-09-982-107-4	Sequence 4, Appl1
7	197	5.6	771	10	US-09-818-247-4	Sequence 4, Appl1
8	186	5.3	769	9	US-09-982-107-10	Sequence 10, Appl1
9	186	5.3	769	10	US-09-818-247-3	Sequence 3, Appl1
10	185	5.3	762	9	US-10-047-542-51	Sequence 51, Appl1
11	183	5.2	732	10	US-09-838-247-5	Sequence 5, Appl1
12	182	5.2	757	9	US-09-982-107-6	Sequence 6, Appl1
13	182	5.2	757	10	US-09-838-247-2	Sequence 2, Appl1
14	175	5.0	1734	10	US-09-862-027-81	Sequence 81, Appl1
15	167	4.8	773	9	US-09-982-107-2	Sequence 2, Appl1
16	167	4.8	773	10	US-09-818-247-6	Sequence 6, Appl1
17	155	4.4	881	10	US-09-816-860-2	Sequence 2, Appl1
18	147	4.2	826	10	US-09-804-908-47	Sequence 47, Appl1
19	145	4.2	1337	10	US-09-803-126-1	Sequence 1, Appl1
20	142	4.1	865	9	US-09-957-958A-19	Sequence 19, Appl1
21	142	4.1	1336	9	US-09-945-901-58	Sequence 58, Appl1
22	141	4.1	1336	9	US-10-007-747-58	Sequence 58, Appl1
23	141	4.1	522	9	US-09-764-668-1138	Sequence 1138, Appl1
24	141	4.1	524	9	US-09-764-668-761	Sequence 761, Appl1
25	141	4.0	332	9	US-09-978-295A-216	Sequence 216, Appl1
26	141	4.0	332	9	US-09-992-598-517	Sequence 517, Appl1
27	141	4.0	332	9	US-09-978-697-216	Sequence 216, Appl1
28	141	4.0	332	10	US-09-989-722-517	Sequence 517, Appl1
29	141	4.0	332	10	US-09-986-723-517	Sequence 517, Appl1
30	141	4.0	332	10	US-09-991-163-517	Sequence 517, Appl1
31	141	4.0	332	10	US-09-989-727-517	Sequence 517, Appl1
32	141	4.0	332	10	US-09-989-731-517	Sequence 517, Appl1
33	141	4.0	332	10	US-09-989-732-517	Sequence 517, Appl1
34	141	4.0	332	10	US-09-991-073-517	Sequence 517, Appl1
35	141	4.0	332	10	US-09-990-442-517	Sequence 517, Appl1
36	141	4.0	332	10	US-09-991-604-517	Sequence 517, Appl1
37	141	4.0	332	10	US-09-990-456-517	Sequence 517, Appl1
38	141	4.0	332	10	US-09-989-721-517	Sequence 517, Appl1
39	141	4.0	338	12	US-10-023-523-43	Sequence 43, Appl1
40	141	4.0	538	12	US-10-023-523-43	Sequence 43, Appl1
41	141	4.0	538	12	US-09-922-171-1068	Sequence 1068, Appl1
42	140	4.0	5179	10	US-09-854-1263-1068	Sequence 1068, Appl1
43	140	4.0	1334	10	US-09-854-173A-12	Sequence 12, Appl1
44	139	3.9	947	10	US-09-871-889-1	Sequence 1, Appl1

```

1 APPLICATION NUMBER: 09/095,385
2 FILING DATE: <unknown>
3 ATTORNEY/AGENT INFORMATION:
4
5 NAME: Canada, Karen S
6 REGISTRATION NUMBER: 39, 927
7 REFERENCE/DOCKET NUMBER: 30435,450S01
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: 310 445-1140
10 TELEFAX: 310 445-9031
11 TELEX: <unknown>
12 INFORMATION FOR SEQ ID NO: 4:
13
14 SEQUENCE CHARACTERISTICS:
15
16 LENGTH: 608 amino acids
17 TYPE: amino acid
18 STRANDEDNESS: unknown
19 TOPOLOGY: linear
20 MOLECULE TYPE: protein
21 SEQUENCE DESCRIPTION: SEQ ID NO: 4:
22 DS-05-950-294-4

```















```

; LENGTH: 602
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-047-542-51

```

Alignment Scores:  
Pred No.:

Pred. No.:	1.69e-05	Length:	612
Score:	185.00	Matches:	115
Percent Similarity:	34.87%	Conservative:	74
Host local Similarity:	21.22%	Mismatches:	186
Query Match:	5.308	Indels:	167
DB:	9	Gaps:	20

US-09-651-150B-1 (1-1911) X US-10-047-542-51 (1-602)

[illegible][illegible]





```

QY 161 -----GGGCGATGAGTAAATGAAAGGCGACTGCT-----GAAAG 199
Db 30 SerValAluGlyArgSerValSerLeuLysSerTyrTyrProThrSerValAsnArg 49
QY 200 CAGTCAGGATATATATGCTGGGAGATGCGTGGATGCGAACATGCTGACGGTGA 259
Db 50 HisThrArgLysTyrTrpCysArgGlnGlyAla---GlnGlyArgSerThrLeu 68
QY 260 TCACACACCAATTCATCAAGGACAAATACAGGCGGAGATTACTGTGAACATACCA 319
Db 69 SerSerGlnGlyTyrValSerAspSerTyrValGlyArgAlaAsnLeuThrAsnPhePro 88
QY 320 CGCAAGATCTGCTGCTAGTGGAGCTAACACACCTGACAGCAAGCTGACACCGAGCTAT 379
Db 89 GluSerGlyThrPheValValAspLeuSerHisLeuThrHisLysAspSerGlyAlaGly 108
QY 380 GCTTGGCCAGCTGGCCATGACACACAGCCGGGAAAGCCAGCAAGCTCACCTGAATGTC 439
Db 109 LysCysGlyLeuGlyIleSerSer ArgGlyLeuAsnPheAspValSerLeuGluVal 127
QY 440 CACAGTCAA 448
Db 128 SerClnAsp 139

RESULT 14
US-09-862-027-81
: Sequence 81, Application US/09862027
: Patent No. US20020142428A1
: GENERAL INFORMATION:
: APPLICANT: Hodge, Martin R.
: TITLE OF INVENTION: NO. US20020142428A1 Kinases and Uses Thereof
: FILE REFERENCE: 35800/234862
: CURRENT APPLICATION NUMBER: US/09/862,027
: CURRENT FILING DATE: 2001-05-21
: PRIOR APPLICATION NUMBER: US 09/345,473
: PRIOR FILING DATE: 1999-06-30
: NUMBER OF SEQ ID NOS: 82
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 81
: LENGTH: 1734
: TYPE: PRT
: OKMISM: Mus musculus
: US-09-862-027-81

Alignment Scores:
Score: 8.49e-05 Length: 1734
Percent Similarity: 32.80% Matches: 137
Host Local Similarity: 24.42% Conservative: 47
Query Match: 5.03% Mismatches: 186
DB: 10 Gaps: 28

US-09-651-150b-1 (1-1911) x US-09-862-027-81 (1-1734)
QY 370 CGGATCTATGCTGGGAGCGGAGATGACACAGACCGGAGAAACCCGAAAGCTAC 429
Db 1191 ArgSerLeu-----SerSerGlyLysSerGlyProGlySerProThrHisLeu 1207
QY 430 -----CGCAATGCTGCCAGACAGCAATACGAGGCAATACGGAAGACGACCAATGCC 480
Db 1208 SerLeuSerProArgSerProPro----- 1215
QY 481 TGAAATTTAAATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 540
Db 1216 -----GlnGly 1217
QY 541 TTCGAATTTCGTAACGACAGCTTACACACC-----ACCTCAAAAGCGGCAAGCT 588
Db 1218 TTTATGVALAlaProAspAlaValHisSerValGlyLysAsnSerGlnSerSer 1237
QY 589 CCGTCGACATCA-----CGACGCTGCGCGCCACCCCAAAATACCCGACGCGCTGCTG 645
||||| 111 111 ||| |||||

```

```

Db 1238 ProSerSerSerValProSerSerProAlaGlySerClnHisThrArgProSerSer 1257
QY 646 CAGAGATTTTCAGTAAAGAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 705
Db 1258 HisGlyLeu-----AlaProLysLeuGlnHisArgSerPro 1271
QY 706 AAAAATTCAGCTCTGGAGGCGGCTGGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAG 762
Db 1272 ArgArgLysSerAlaGlySerLeuProLeuSerTrpLeuAlaHisThrProSerPro 1291
QY 763 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 822
Db 1292 AlaThrAlaAlaSer----- 1296
QY 823 AGGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 882
Db 1297 -----ProLysArgSerProSerProLeuSerGlyHis 1307
QY 883 GCTGCTGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 942
Db 1308 GlySerGln-----SerPheProThrLysLeuHisLeuSerProProLeuGly 1323
QY 943 GGTAGAGATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 993
Db 1324 ArgGlnLeuSerArgProLysSerAlaGluProProArg-----SerProLeuLysArg 1342
QY 994 ACCGAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1024
Db 1342 GVALGlnSerAlaGlyLysLeuAlaAlaAlaLeuAlaAlaGlyLysLeuAlaArg 1362
QY 1025 -----GCGCGCG 1032
Db 1362 SerArgLysHisSerLeuAspLeuProHisGlyGlyLeuLysGlySerLeuThrPro 1382
QY 1033 GAGAGATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1077
Db 1382 GGLuLaserProLeuGluValGlyThrArgSerValLeuSerGlyLysGlyProLeu 1402
QY 1078 TCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137
Db 1402 UProlLysGlyVal-----LeuGlnProAlaPro-----SerArgAlaLe 1416
QY 1138 CAGTCCGCAATCTCGAAGACAGCTGTGAATAGCTAGCTTTAGTAAAGAGAGAGAG 1197
Db 1416 uGlyThrLeuArgGlnAspArgAlaGlnArgArgGlySerLeuGlnLysGlnAlaAla 1436
QY 1198 CATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1245
Db 1436 eArgGluValAspSerSerGlyAspAspHisAspGluLeuProLysAsnSerGlnAla 1456
QY 1245 AGTAAATTTAGTAAATTTAGTAAATTTAGTAAATTTAGTAAATTTAGTAAATTTAG 1278
Db 1456 rGlnLysProAlaGlyLeuSerProHisProHisProHisLeuLeuProLysGlyLe 1476
QY 1279 GCACTGTGCTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1335
Db 1476 rGlyLysLeuThrGlnLysAspThrPheLeuHisArgAspLeuLys--LysGlnGlyPro 1495
QY 1336 GTTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1395
Db 1496 ValLeuSerGlyLeuValThrGly-----AlaThrLeuGlySerProAlaGlyValAsp 1512
QY 1396 GAGAGATTTAGAGATTTAGAGATTTAGAGATTTAGAGATTTAGAGATTTAGAGAT 1428
Db 1513 ValPro-----GlyLeuSerProArgLysValSerArgProGlnAlaPheGlnAla 1530
QY 1429 TCGACACGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488
Db 1531 ThrAsnProLeuGlnValProSerLeuSerArg-----SerGlyPro 1544
QY 1489 CCGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1548
Db 1545 ThrSerProThr-----ProSerGlnGlyCysTrpLysAlaGlnHisGln 1559

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 24, 2002, 05:50:30 ; Search time: 48 seconds

7654.703 Million cell updates/sec

Title	NS - 09-651-150R-1
Trustee	3400

Sequences 1 aaagqaaq1 aaqcaqgq1q1e 11aaatct1g1ct1ccaa1ccct11r 1911

### Scoring table:

Searched: 283224 seqs, 96134422 residues

Total number of hills satisfying chosen parameters: 566448

Maximum DB seq length: 200000000000

post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

```

MODEL=frame,n2p,model=-DBV-X1P
-0/-cgrn,1/0/SPMT,spool/0/5036551153/runaL20112002,130744,23870/3pf-queri.fasta.1.20555
-DB-P1R-73 -OMT-laslan -SOFTX-62p.rpr -MNMATCH-0.1 -LOOPCL-0 -LOOEXT-0
-DUNTS-bits -START-1 -END-1 -MATRIX-blossum2 -TRANS-human45.cdi -LIST-45
-DOCALLIN-200 -THR-STOPE-p-1 -THR-MAX-100 -THE-MIN-0 -ALIN-15 -MODE-LOCAL
-OUPMT-pte -MEMO-ext -HEAPS-ext -MINLEN-0 -MAXLEN-2000000000
-USER-0509451150.accl.1.49.arunat.20112002,100744,23870 -NCGR-6 -ICPD-3
-NO-X1XP -NO-MMP -LACJGUEHRY -NRC SCORES-0 -WAIT -LONGLOG -DEV-TIMEOUT-120
-WARR-TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-10 -5 -FAPAP-6 -FAPAEEXT-7
-YFAPOP-10 -YFAPEXT-0 -5 -DEV-P-6 -DELET-10 -X

```

Database :

```

pir_73.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No	Score	Query Match	Length	DB	ID	Description
1	200	5.7	764	1	Q8R1G3	secretory component
2	186	5.3	764	1	GPTP3S	secretory component
3	182	5.2	757	1	A5484.1	secretory component
4	182	5.2	757	2	I45346	polymeric immunoglobulin receptor
5	180.5	5.2	1184	2	G01763	atrophin-1 - human
6	178	5.1	1184	2	S53812	atrophin-1 - human
7	175.5	5.0	1734	2	A54602	microtubule-associated protein 1 - human
8	173	5.0	4957	2	T03455	microtubule-associated protein 1 - human
9	167.5	4.8	773	1	Q8R1G3	secretory component
10	163	4.7	1357	2	T29265	secretory component
11	162.5	4.7	620	2	S06733	hypothetical protein
12	162	4.6	2715	2	T13849	hydroxymethylglutaryl-CoA lyase
13	161	4.6	1839	1	PPDEP1	yeast fruit fly
14	159	4.6	1224	2	T40765	genome polyprotein

15	159	4.6	1.952	2	148814	hypothetical protein
16	158	4.5	1.298	1	EDME75	immediate-early protein
17	157	4.5	1.283	2	S13863	hydroxyproline-rich glycoprotein
18	156	4.5	1.006	2	142731	atrophin-1-related protein
19	155.5	4.5	3.504	2	A59266	unconventional myosin
20	155	4.4	6.49	2	146500	hypothetical protein
21	155	4.4	8.47	2	F96531	hypothetical protein
22	153	4.4	801	2	T29018	hypothetical protein
23	153	4.4	825	1	EDH8XD	immediate-early protein
24	151	4.3	416	2	J00465	extensin precursor
25	150	4.3	946	2	S27921	nuclear antigen F8
26	149.5	4.3	369	2	S20500	hydroxyproline-rich glycoprotein
27	149	4.3	9.43	2	S52796	PLP2 protein - human
28	149	4.3	9.44	2	S27923	gene LF3 protein - human
29	148.5	4.3	760	2	F86387	probable pro kinase
30	148.5	4.3	817	2	S51342	verpoxin
31	148	4.2	819	2	T04859	extensin homolog F
32	146.5	4.2	3.38	2	J20085	hydroxyproline-rich glycoprotein
33	145.5	4.2	1.036	2	T03469	hydroxyproline-rich glycoprotein
34	145.5	4.2	16.11	2	T38336	hypothetical protein
35	145.5	4.1	350	2	S22456	hydroxyproline-rich glycoprotein
36	143.5	4.1	13.30	1	D23018	collagen alpha 4(I)
37	143	4.1	10.67	2	A55617	mucinase precursor
38	143	4.1	30.20	2	A43932	mucin 2 precursor
39	142.5	4.1	1.733	2	S27939	lensin - chicken
40	142	4.1	399	2	S10899	proline-rich protein
41	142	4.1	990	2	T14756	hypothetical protein
42	142	4.1	1560	2	T00080	hypothetical protein
43	142	4.1	2187	2	T30826	nascent polypeptide
44	141.5	4.1	4.98	2	T24759	hypothetical protein
45	141.5	4.1	577	2	T09024	proline-rich protein

## ALIGNMENT'S

**RESULT 1**

**ORHOGS**

secretory component precursor [validated] - human

N.Alternate names: poly Ig receptor, polymeric immunoglobulin receptor

N.Containers: free secretory component, transmembrane secretory component

C.Species: Homo sapiens (man)

C.Date: 28-Aug-1965 #sequence\_revision 24-Apr-1996 #text\_change 08-Dec-2000

C.Accession: A46537; A55284; I38115; A32263; S38978; S13453; A02112

R.Krajal, P.; Kvale, D.; Tasken, K.; Brandtzaeg, P.

Pur.: J. Immunol., 22, 2369-2375, 1992

A.Title: Molecular cloning and exon-intron mapping of the gene encoding human transmembrane secretory component (poly-Ig receptor); molecular c

A.Reference number: A46537; MUID:92487236; PMID:1355431

A.Accession: A46537

A.Status: not compared with conceptual translation

A.Molecule type: DNA

A.Residues: 1-764 <KRA>

A.Cross-references: CH 543449; NID:3255047; FIDN AAH23776.1; PID:9255098

A.Experimental source: leukocytes

A.Note: sequence extracted from NCBI backbone (NCBIP:113253)

R.K.Rajic, E., Grzeschik, K.H., Geurts van Kessel, A.H., Gidsen, B.; Brandtzaeg, P.

Jum. Genet., 87, 642-648, 1991

A>Title: The human transmembrane secretory component (poly-Ig receptor); molecular c

A.Reference number: A55284; MUID:92039621; PMID:1682231

A.Accession: A55284

A.Molecule type: mRNA

A.Residues: 1-764 <KR>

A.Cross-references: ND 562403; NID:9238235; FIEN AAH55284.1; FID:3238236

A.Experimental source: colonic adenocarcinoma cell line

A.Note: sequence extracted from NCBI backbone (NCBIN:62403; NCBIP:62408)

R.Piskurich, J.F.; France, J.A.; Tamer, C.M.; Willmer, C.A.; Kaetzl, C.S.; Kaetzl,

MOL. Immunol., 30, 413-421, 1993

A>Title: Interferon-gamma induces polymeric immunoglobulin receptor mRNA in human int

A.Reference number: I38115; MUID:93205018; PMID:8455639

A.Accession: I38115

A.Molecule type: mRNA

A.Residues: 1-764 <RES>

A.Cross-references: EMBL:X73079; NID:9456345; FIDN CMA1522.1; FID:9456346

A>Note: submitted to the EMBL/Genbank/DBJ databases by J.F. Piskurich, February 1994













```

QY 823 AGCATTCACATCTGATCCGACCAACCTGCGCCCTTTCTGCTGCGACCTTCGCGGCT 882
Db 1297 -----ProGlnArgSerProSerProSerGlyHis 1307
QY 883 GGTGATGAAAAAGGATTTTAAAGAGAAAGATTTTAAAGAGAGAGAGAGAGAGAGAGAG 942
Db 1308 GlySerGln-----SerProProThrHisLeuHisLeuSerProLeuGly 1323
QY 943 CATTAAATATATATATATATATATATATATATATATATATATATATATATATATATAT 993
Db 1324 ArgGlnLeuSerArgPro-LysSerLacIuProProGln-SerProLeuLeuLysArg 1342
QY 994 ACCCGCCCTCCCAACACACATCTACACGCCCC----- 1024
Db 1342 GValGlnSerAlaGluLysLeuAlaAlaAlaAlaAlaAlaGluLysLysLeuAlaArg 1362
QY 1025 -----TCCCGCGC 1032
Db 1362 GSerArgLysHisSerLeuAspLeuProHisLysGluLeuLysGluLeuThrProArg 1382
QY 1033 GTCGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1077
Db 1382 GValAlaSerProLeuGluValValGlyThrArgSerValLeuSerGlyLysGlyProLeu 1402
QY 1078 TCCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCC 1137
Db 1402 uProGlyLysGlyVal-LysGlnProAlaPro-----SerArgAlaLeu 1416
QY 1138 GCAATGTCATTTTAAATATATATATATATATATATATATATATATATATATATATATAT 1197
Db 1416 uGlyThrLeuArgGlnAspArgAlaGluArgArgGluSerLeuGlnLysGlnAlaLeu 1436
QY 1198 CATGATGGAGAGAGAGATATATATATATATATATATATATATATATATATATATATATAT 1245
Db 1436 eArgGlnValAspSerSerGluAspSerThrAspGluGluProGluLysSerGlnAlaLeu 1456
QY 1246 ACAAGTCCGACCATATCCGCAACCCCA-----GCCGC 1278
Db 1456 rGlnGluProArgLeuSerProHisProGluAlaSerHisAsnLeuLeuProLysGlySe 1476
QY 1279 GGAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1335
Db 1476 rGlyGluGlyThrGluLysAspThrLeuHisArgAspLeuLys--LysGlnGlyPro 1495
QY 1336 GTTCTAGACCCCTCATACATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1395
Db 1496 ValLeuSerGlyLeuValThrGly-----AlaThrLeuGlySerProArgValAsp 1512
QY 1396 GCGCGTACGATGCGCTCGCCCGCAGC-----TCGCT 1428
Db 1513 ValPro-----GlyLeuSerProArgLysValSerArgProGlnAlaPheGlnLysAla 1530
QY 1429 TGCACACCTTGGACAGATATATATATATATATATATATATATATATATATATATATATAT 1488
Db 1531 ThrAsnProLeuGlnValProSerLeuSerArg-----SerGlyPro 1544
QY 1489 CCAATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1548
Db 1545 ThrSerProThr-----ProSerGluGlyCysTrpLysAlaGlnHisLeu 1559
QY 1549 TCTTCGATTCGCTTCGAC-----ACAGCGCT 1575
Db 1560 HisThrGlnAlaLeuThrAlaLeuCysProSerPheSerGlyLeuThrProThrGlyCys 1579
QY 1576 TGTATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1614
Db 1580 SerAlaAlaThrSerThrSerGlyLysProGlyThrTrpSerTrpLysPheLeuLysGlu 1599
QY 1615 GATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1674
Db 1600 GlyProAspArgAlaSerThrAsnLysThrLeuThrArgLysGlyLysProLysAsnSer 1619
QY 1675 AGC-----CCTTCGCTCGCTGATATAT-----GACCT 1701

```

```

Db 1620 GlnAspThrAsnThrThrValProAsnGlnLeuLysAsnGlnSerProGlnGluGly 1639
QY 1702 GATTCAGATATATATATATATATATATATATATATATATATATATATATATATATATAT 1756
Db 1639 SPFGInProProSerValPro---GlyLeuThrHisProLeuGluValPro 1656
RESULT 8
T03455
ALR protein - human
C:Species: Homo sapiens (man)
C:Date: 24 Mar 1999 #sequence_revision 24-Mar-1999 #text_change 21 Jul 2000
C:Accession: T03455
R:Prasad, R.; Zhadanov, A.R.; Sedkov, Y.; Bullrich, F.; Bruck, T.; Kallipatla, R.; Ya
Oncogene 15, 549-560, 1997
A:Title: Structure and expression pattern of human ALR, a novel gene with strong homo
A:Reference number: Z14954; M01D:97388474; EMBID:9247308
A:Accession: T03455
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4957 (FEA)
A:Cross-references: EMBL:AF010404; NID:92358286; FIDN:AA051735.1; FID:92350287
A:Genes:
A:Gene: ALR
A:Map position: 12
A:Superfamily: human ALR protein
C:Keywords: alternative splicing
Alignment Scores:
Pred. No. 0 000323 length: 4957
Score: 173.00 Matches: 149
Percent Similarity: 31.98% Conservative: 48
Best Local Similarity: 24.19% Mismatches: 174
Query Match: 4.96% Indels: 245
Gaps: 34
US-09-651-150b-1 (1-1911) x T03455 (1-4957)
QY 452 GAGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 311
Db 99 ProThrSerProProProGlnAlaSerArgLeuSer-ProProProGlnAspSerProThr 118
QY 312 AATACCCAGCCAGCAATCTGCTCCCTACGCGAGGTACACACAGCTGACAGA----- 360
Db 118 rSerProProProGlnAspSerProAlaSerProProProGlnAspSerLeuMetSerLe 138
QY 361 -----AGTGAACAGAGATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 389
Db 138 uProLeuGluLysSerProLeuLeuProLeuProGluGluProGluGlnCysProArgSe 158
QY 390 GAGGATGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 435
Db 158 rGluGlyProHisLeuSerProArgProGluGluProHisLeuSerProArgProGluGlu 178
QY 436 -----TCGTCACATGATATGACGACATGACGACATGACGACATGACGACATGACGACAT 475
Db 178 uProHisLeuSerProGln-AlaGluGluProHisLeuSerProGlnProGluGluPro 198
QY 476 -----ATGCTGAGAGATGCAAAATGCTTATG-----C 506
Db 198 LysLeuGlnAlaValProGluGluPro-----HisLeuSerProGlnAlaGluGlyP 215
QY 507 GCAATTTGTCACATATATGATATATGATATATGATATATGATATATGATATATGATATATGAT 566
Db 215 rHisLeuSerProGlnProGluGluLeu----- 224
QY 567 CACCAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 626
Db 225 -----HisLeuSerProGlnProGlnProGlnProGlnProGlnProGlnProGlnPro 233
QY 627 GCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 686
Db 233 rHisLeuSerProVal-ProGluGlu-----ProCysLeuSerProGlnProGlu 249

```









Query Match:	4.618	Indels:	169
DP:	1	Gaps:	28

[illegible][illegible]

## RESULTS 14

web1 protein homolog - fission yeast (Schizosaccharomyces pombe)  
 C.Species: Schizosaccharomyces pombe  
 C.Date: 03-dec-1999 #sequence revision 03-dec-1999 #text\_change 03-dec-1999  
 C.Accession: J40765  
 R.Lyte, M.; Ralandrac, M.A.; Barrell, B.G.; Jaber, J., Hilbert, H., Dusterhoeft, A.  
 submitted to the EMBL Data Library, March 1998  
 A.Reference number: Z21948  
 A.Accession: J40765  
 A.Status: preliminary: translated from CH/EMBL/DDBJ  
 A.Molecule type: DNA  
 A.Residues: 1-1224 <1>N>  
 A.Cross-References: EMBL AL022072, FIDN FNA1844, 1: 6SFEB, GUN0647, SFEB, SPBC8D2, 20C  
 A.Experiment: strain 972h-; cosmid c8D2  
 C.Genetics:  
 A.Gene: SPDB:SPBC8D2, 20C  
 A.Map position: 2

### Alignment Scores:

Pred. No.:	0	0024	Length:	1234
Score:	159	00	Matches:	126
Percent Similarity:	32.118		Conservative:	58
Best local Similarity:	21.994		Mismatches:	218
Query Match:	4.568		Indels:	172
DB:	2		Gaps:	21

US-09-651-150B-1 (1-1911) X T40765 (1-1224)

O<sub>7</sub> 287 TACAAAGCGCCAGTTACTCTGACGCAATACCCAGCGCAACATCITTCCTACATGACGCTA 346  
 ||| ..... ||| |||||.....  
 Db 679 TyrtleAlaserLysSerleuGlnSerTyr...AlasnlLeuTrrpleuLysGlnleu 696



























FT	CHAIN	19	773	POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
FT	DOMAIN	19	615	SECRETORY COMPONENT.
FT	DOMAIN	19	647	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	648	670	POTENTIAL.
FT	DOMAIN	671	773	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	39	122	IG-LIKE V-TYPE DOMAIN 1.
FT	DOMAIN	148	232	IG-LIKE V-TYPE DOMAIN 2.
FT	DOMAIN	253	331	IG-LIKE V-TYPE DOMAIN 3.
FT	DOMAIN	362	445	IG-LIKE V-TYPE DOMAIN 4.
FT	DOMAIN	471	545	IG-LIKE V-TYPE DOMAIN 5.
FT	DISULFID	46	115	POTENTIAL.
FT	DISULFID	155	225	POTENTIAL.
FT	DISULFID	260	324	POTENTIAL.
FT	DISULFID	369	438	POTENTIAL.
FT	DISULFID	478	538	POTENTIAL.
FT	CARBOHYD	88	88	POTENTIAL.
FT	CARBOHYD	108	108	POTENTIAL.
FT	CARBOHYD	418	418	POTENTIAL.
FT	VARIANT	88	94	N-LINKED (GLCNAC. . .); IN ALLOTYPE T62
FT	VARIANT	94	94	K > N (IN ALLOTYPE T61).
FT	VARIANT	101	108	D > E (IN ALLOTYPE T61).
FT	VARIANT	110	110	IVDULFVN -> YLNLSLS (IN ALLOTYPE T61).
FT	VARIANT	110	110	S > T (IN ALLOTYPE T63).
FT	SEQUENCE	773 AA.	81886 MW,	DP24142F11Y4655 JK664.

US-09-651-150H-1 (1-1911) x P1CR_RAB11 (1-773)	Score:	0.10105	length:	773
Percent Similarity:	167.50		Matches:	52
Best Local Similarity:	53.85%		Conservative:	18
Query Match:	4.00%		Mismatches:	46
DB:	1		Indels:	14
			Gaps:	6

Alignment Scores:	pred. No.:	0.10105	length:	773
Score:	167.50		Matches:	52
Percent Similarity:	53.85%		Conservative:	18
Best Local Similarity:	4.00%		Mismatches:	46
Query Match:	4.00%		Indels:	14
DB:	1		Gaps:	6

US-09-651-150H-1 (1-1911) x P1CR_RAB11 (1-773)	Score:	0.10105	length:	773
Percent Similarity:	167.50		Matches:	52
Best Local Similarity:	53.85%		Conservative:	18
Query Match:	4.00%		Mismatches:	46
DB:	1		Indels:	14
			Gaps:	6

US-09-651-150H-1 (1-1911) x P1CR_RAB11 (1-773)	Score:	0.10105	length:	773
Percent Similarity:	167.50		Matches:	52
Best Local Similarity:	53.85%		Conservative:	18
Query Match:	4.00%		Mismatches:	46
DB:	1		Indels:	14
			Gaps:	6

US-09-651-150H-1 (1-1911) x P1CR_RAB11 (1-773)	Score:	0.10105	length:	773
Percent Similarity:	167.50		Matches:	52
Best Local Similarity:	53.85%		Conservative:	18
Query Match:	4.00%		Mismatches:	46
DB:	1		Indels:	14
			Gaps:	6

US-09-651-150H-1 (1-1911) x P1CR_RAB11 (1-773)	Score:	0.10105	length:	773
Percent Similarity:	167.50		Matches:	52
Best Local Similarity:	53.85%		Conservative:	18
Query Match:	4.00%		Mismatches:	46
DB:	1		Indels:	14
			Gaps:	6

US-09-651-150H-1 (1-1911) x P1CR_RAB11 (1-773)	Score:	0.10105	length:	773
Percent Similarity:	167.50		Matches:	52
Best Local Similarity:	53.85%		Conservative:	18
Query Match:	4.00%		Mismatches:	46
DB:	1		Indels:	14
			Gaps:	6

US-09-651-150H-1 (1-1911) x P1CR_RAB11 (1-773)	Score:	0.10105	length:	773
Percent Similarity:	167.50		Matches:	52
Best Local Similarity:	53.85%		Conservative:	18
Query Match:	4.00%		Mismatches:	46
DB:	1		Indels:	14
			Gaps:	6

US-09-651-150H-1 (1-1911) x P1CR_RAB11 (1-773)	Score:	0.10105	length:	773
Percent Similarity:	167.50		Matches:	52
Best Local Similarity:	53.85%		Conservative:	18
Query Match:	4.00%		Mismatches:	46
DB:	1		Indels:	





DB 842 TyrosyllealaserSerSerThrValIleHisTyrInProGlyPro----- 859  
 QY 1393 TGTGGCGCTGTAGCAAT-----CCGTCGTGCCCCCACTGCTCT 1428  
 DB 860 -----PRTPTVHISPhSerSerLeuSerProArgLeuAlaIaIaSerAlaDProArg----- 875  
 QY 1429 TGCACACCT 1437  
 DB 876 CysasnPro 878  
 RESULT 11  
 SHKL RAT  
 ID SHKL RAT STANDARD: PPT: 2167 AA  
 AC Q9WV48, Q9WV13, Q9WTER, Q9GZ28,  
 DT 15-JUN-2002 (Ref. 41, Created)  
 DT 15-JUN-2002 (Ref. 41, last sequence update)  
 DE 15-JUN-2002 (Ref. 41, last annotation update)  
 DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1) (GKAP/SAPAP  
 DE Interacting protein) (SSTR Interacting protein) (SSTRIP).  
 GN SHANK1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID-10116;  
 RX [1]  
 RP SHONCE: FROM N A (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND  
 DI04.  
 RC TISSUE-Brain;  
 RA MEDLINE-99419021; PubMed-10488079;  
 RX Yao I., Hala Y., Hira K., Deguchi M., Ide N., Takeuchi M., Takai Y.;  
 PT "Synanon, a novel neuronal protein interacting with synapse-associated  
 RT protein 90/postsynaptic density-95-associated protein.";  
 RL J. Biol. Chem. 274:27463-27466(1999).  
 RN [2]  
 RP SEQUENCE FROM N A (ISOFORM 4), AND INTERACTION WITH DLGAP1.  
 RC STRAIN-Sprague-Dawley;  
 RA MEDLINE-99350650; PubMed-10433268;  
 RX Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtchanoff J.,  
 RA Weinberg R.J., Worley P.F., Sheng M.;  
 RT "Shank, a novel family of postsynaptic density proteins that binds to  
 RL the NMDA receptor/PSD-95/GKAP complex and cortactin.";  
 RN Neuron 23:569-582(1999).  
 RN [3]  
 RP SEQUENCE FROM N A (ISOFORM 1).  
 RC TISSUE-Brain;  
 RA MEDLINE-20549637; PubMed-10958799;  
 RX Tobaben S., Suedhof T.C., Stahl B.;  
 RT "The G protein-coupled receptor CIL interacts directly with proteins  
 RL of the Shank family.";  
 RN J. Biol. Chem. 274:29510-29518(1999).  
 RN [4]  
 RP PARTIAL SEQUENCE FROM N A (ISOFORMS 1; 2; 3; 4 AND 5), AND  
 RP DEVELOPMENTAL STAGE.  
 RC TISSUE-Brain;  
 RA MEDLINE-99436166; PubMed-10506216;  
 RX Lim S., Naisbitt S., Yoon J., Huang J.L., Suh P.G., Sheng M., Kim E.;  
 RT "Characterization of the Shank family of synaptic proteins. Multiple  
 RT genes, alternative splicing, and differential expression in brain and  
 RL development.";  
 RN J. Biol. Chem. 274:29510-29518(1999).  
 RN [5]  
 RP PARTIAL SEQUENCE FROM N A (ISOFORM 4).  
 RC TISSUE-Brain;  
 RA MEDLINE-20020275; PubMed-10551867;  
 RX Zilzer H., Hennek H.-B., Baechner D., Richter D., Kretenkamp H.-J.;  
 RT "Somatostatin receptor interacting protein defines a novel family of  
 RT multidomain proteins present in human and rodent brain.";  
 RL J. Biol. Chem. 274:32997-33001(1999).  
 RN [6]  
 RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.  
 RX PubMed-10433269;  
 RA Tu J.C., Xiao B., Naisbitt S., Yuan T.P., Petralia R.S., Brakeman P.,

RA Doan A., Akalu V.K., Lanahan A.A., Sheng M., Worley P.F.;  
 RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of  
 RT postsynaptic density proteins.";  
 RL Neuron 23:583-592(1999).  
 RN [7]  
 RP INTERACTION WITH SEPTIN.  
 RX PubMed-11509555;  
 RA Bockers T.M., Mameza M.G., Kreutz M.F., Bockmann J., Weise C.,  
 RA Buck P., Richter D., Gundelfinger E.D., Kretenkamp H.-J.;  
 RT "Synaptic scaffolding proteins in rat brain. Ankyrin repeats of the  
 RT multidomain Shank protein family interact with the cytoskeletal  
 RL protein alpha-todrin.";  
 RN J. Biol. Chem. 276:40104-40112(2001).  
 RN [8]  
 RP FUNCTION.  
 RX PubMed-11498055;  
 RA Sala C., Piech V., Wilson N.R., Passafium M., Liu G., Sheng M.;  
 RT "Regulation of dendritic spine morphology and synaptic function by  
 RT Shank and Homer.";  
 RL Neuron 31:115-130(2001).  
 RN [9]  
 RP REVIEW.  
 RX PubMed-10806096;  
 RA Sheng M., Kim E.;  
 RT "The Shank family of scaffold proteins.";  
 RL J. Cell Sci. 113:1851-1856(2000).  
 CC -1- FUNCTION: Seems to be an adapter protein in the postsynaptic  
 CC density (PSD) of excitatory synapses that interconnects receptors  
 CC of the postsynaptic membrane including NMDA-type and metabotropic  
 CC glutamate receptors, and the actin-based cytoskeleton. May play a  
 CC role in the structural and functional organization of the  
 CC dendritic spine and synaptic junction. Overexpression promotes  
 CC maturation of dendritic spines and the enlargement of spine heads  
 CC via its ability to recruit Homer to postsynaptic sites, and  
 CC enhances presynaptic function.  
 CC -1- SUBUNIT: May homodimerize via its SAM domain. Interacts with  
 CC SPAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with  
 CC DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via  
 CC the PDZ domain (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic, postsynaptic density of  
 CC neuronal cells. Colocalizes with alpha-latrotoxin receptor 1,  
 CC 4/a and 5; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed only in brain (neuropil of cortex,  
 CC CA1 region hippocampus and molecular layer of cerebellum).  
 CC DEVELOPMENTAL STAGE: Expression increases from low levels at birth  
 CC to high levels at 3-4 weeks before dropping slightly in adulthood.  
 CC Expressed in the cortex and the molecular layer of the cerebellum  
 CC at postnatal day 7. Isoform 2 expression does not change during  
 CC development of both cortex and cerebellum. Isoform 4 expression  
 CC decreases significantly during development of cortex but not  
 CC cerebellum.  
 CC -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.  
 CC -1- SIMILARITY: CONTAINS 7 ANK REPEATS.  
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AF102855; AAD04569.2; -;  
 DR EMBL: AF131951; AAD29617.1; ALT\_INIT.  
 DR EMBL: AF159046; AAD42975.1; -;  
 DR EMBL: AF141904; AAF02498.1; ALT\_INIT.  
 DR HSSP: P00519; IABI.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR001478; PDZ.





D6		r11alialarctgysaliaglyproaliproaliaia-----Glnalialiacysl	884
OY	1356	TTCATTTCTGGTACATTGTAAATCCCA-----TGATTTGGG	1397
D6	884	yglvglvAspaSpsasPaspnProhlssqIyalaiagiylArGLephecl	904
OY	1398	cCtIGAGAcAtGCCcGCcccccCACAgCGICGHCACACCCTGCACD-----CCCC	1448
D6	904	yProlenAralaserltyProlenuRgaGMelAlalaiTPMeLArgInIllePrGs	924
OY	1449	TGTAgtGaCaGGAAGCTGTatGCATGTAGAATAITCCC-----	1491
L6	924	pTroCluaSpvalARgyVaIVaIIValyISerTerroLeuprociLyGuASplacuaiagi	944
OY	1492	-----ANNGCANCTTGCTT	1505
D6	944	yGlvgLIvaIserylgltyProProcultIPterSerAlagiuaqglyLeulerCySe	964
OY	1506	CQTTCACAAGCcGTGACACAGACACTGCAGATTGCAGAGCHGTCTCATCACTCTTAGC	1565
D6	964	uleualalaLeualaSnaruLeucycsglyProaspHrilaialiatprlaglyASnTr	984
OY	1566	GAnAsptTGtttgTggTCGMAApTpTaATATATATAGATATAGATAGATAGAT	1625
L6	984	pPhrcly-----AlaprosPalSerAlaleucily--AlaIngliyva	998
OY	1626	AGCATTTGCTCG-----GGCATCTT	1648
D6	998	IleuleuenserThrargrspLeulaIAphealaglYaIaGIuephelyeulyLeute	1018
OY	1647	TCCAGAGTTCOCCTCTTACACAAATAGAACGCTCTTCCTCACTTATGTACGTCCTC	1706
D6	1018	uAlaseRalady-----ASPdaIggleullelaivalIASnrThvalARGAd-	1034
OY	1707	AGCCCATGAGATAGACAGGGGCTGTGTATAAACATCTTGGAAAAGCCCTTGGCTTAT	1766
D6	1035	-CysaspIprioAIaspjolyprioAlavalserargin---HisalatytIaeuadiy	1053
OY	1767	CCAATGITA 1776	
D6	1053	sGUleudeu 1056	
<b>RESULT 13</b>			
<b>TRX2_HUMAN</b>			
ID	TRX2_HUMAN	STANDARD:	FPT: 2715 AA.
AZ	VQWMMK_29UKxLS_495836_vqvvr_q.vqwmmk_115-22_creatid_y96fP3;		
P1	16-OCT-2001 (Ref. 40, Created)		
P2	15-JUN-2001 (Ref. 40, Last sequence update)		
P3	15-OCT-2002 (Ref. 41, Last annotation update)		
D6	TriThorax homolog 2 (Mixed lineage leukemia gene homolog 2 protein).		
NB	TRX2 or HFX2 or MLF2 or MLL4 OR KIA00304.		
OS	Homo sapiens (human).		
NC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;		
JC	Mammalia; eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (LONG ISOFORM).		
RA	Angrand P G., Valvatne H., Jeanmougin F., Adams A.,		
RA	van der Hoeven F., Olsen L., Tekotte H., Huang N., Poeh O.,		
RA	Janovic J., Chambon P., Lossos R., Stewart A., Asland K.,		
K1	"Mamalian TriThorax and ASH1-like proteins, putative chromatin		
K1	regulators which contain PHD fingers and SH domains.";		
RL	Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.		
FN	[2]		
FP	SEQUENCE FROM N.A. (LONG ISOFORM).		
RA	Imertin J.E., McCreedy P.M., Adams A.W., Burkhardt-Schultz K.,		
RA	Garcia E., Kyle A., Ramirez M., Stillwagen S., Garnes J., Dangnan L.,		
RA	Bruce R., Quan Q., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,		
RA	Carrau A.V.;		
RT	*Sequence analysis of a 1 kb region in human jql3.1.*;		
SL	Submitted (NOV-1996) to the EMBL/Genbank/DDBI databases.		
RN	[3]		
RF	SEQUENCE OF 11-2715 FROM N.A. (LONG ISOFORM).		







DR SMART: SM00015; 10; 2.  
 DR SMART: SM000342; MYSH: 1.  
 DR SMART: SM00139; MYTH4; 2.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS00660; BAND\_41\_1; FALSE\_NEG.  
 DR PROSITE: PS00661; BAND\_41\_2; FALSE\_NEG.  
 DR PROSITE: PS50057; BAND\_41\_3; 1.  
 DR PROSITE: PS50096; 10; 3.  
 DR PROSITE: PS50002; SH3; FALSE\_NEG.  
 DR MYOSIN: ATP-binding; Actin-binding; Repeats; SH3 domain;  
 DR Calmodulin-binding; Disease mutation; Deafness;  
 KM DOMAIN 1 1887 HEAD OR MOTOR DOMAIN.  
 F1 DOMAIN 1888 2029 NECK OR REGULATORY DOMAIN.  
 F1 DOMAIN 2030 3530 TAIL.  
 FT DOMAIN 1323 1350 COILED COIL (POTENTIAL).  
 FT DOMAIN 1792 1799 ACTIN-BINDING (POTENTIAL).  
 FT DOMAIN 1902 1924 10 1.  
 FT DOMAIN 1925 1954 10 2.  
 F1 DOMAIN 1955 1976 10 3.  
 F1 DOMAIN 2867 2953 SH1.  
 FT DOMAIN 3206 3443 BAND 4.1-LIKE.  
 FT NP BIND 1315 1322 N -> Y (IN DENB3); FAMLY FROM BENGALA).  
 FT VARIANT 2111 2111 /FTID-VAR-010303.  
 F1 VARIANT 2113 2113 /FTID-VAR-010304.  
 F1 SEQUENCE 3530 AA: 395171 MW: 3D103923D4BCBEAA CRC64:  
 SU  
 Alignment Scores:  
 Pred. No: 0.0527 Length: 3530  
 Score: 155.50 Matches: 108  
 Percent Similarity: 30.13% Conservative: 32  
 Best local Similarity: 23.23% Mismatches: 161  
 Query Match: 4.46% Indels: 164  
 DB: 1 Gaps: 22  
 US-09-651-150b-1 (1-1911) x MY15\_HUMAN (1-3530)  
 QY 274 CATCAACAGCAATATVAAAGAGAGAGATTAATTTGAAGCAATACCAAG----- 321  
 DB 552 HISAAGQLYLEUGLYPHEGLYPROGLPHEGLYARGPROVALPRORARPROVALTHRSE 571  
 QY 321 ----- 321  
 DB 572 LEUALAARQPHLEULYSTYSTRLEUSERGLULYLSYSPROILLEAARQLEUARGLY 591  
 QY 322 CAACAAATCTTCTTACTGAGAGATTAACAGAGTAACAGAAAGTGAAGGAGATCTATG 381  
 DB 592 SERGLNYSITRARGALAGLYGLYPROVALAVAL-----ARGGLUALAALATYR 607  
 QY 382 CTGGATACAGAGATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 441  
 DB 608 LYSARQPHGLYGLYFLYS-----LEUALAGLYMELASPROGLULYS---PROGLYTHPR 625  
 QY 442 CAGTCAATATGAGCAATCAICAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 501  
 DB 626 ILEVALLEUARGAVALAGLAPROVALARGSERSETASNPALARGARQPROPO 645  
 QY 502 TGTATCTATTGTTTAAATATCTGATATGAGAGATCTTAAATGAGAGAGAGAGAG 561  
 DB 646 ALARPHROGLNPRVALPR----- 651  
 QY 562 TACCAATATTAATTAAG 621  
 DB 652 ---ARQTHLEUSEHISTIPSERALALEULEUSERPROVALPRORARPROPO 670  
 QY 622 AATGACCCAGCCCTCTGACATGACATGACATGACATGACATGACATGACATGACATGAC 681  
 DB 671 SERSERGLYPROPO-----PROVALPRORARPROPO 679  
 QY 682 CTCTCTCTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 735  
 DB 680 LEUSERPROVALLEUSERGLYGLYPRORARPROVALASERTRIOYLYGLYSERLEUARG 699

QY 736 GCCCCAGACGCGCCAGCTACAACTACACACTAGAGCTGCTACAGAGAGAGAGAGAGAG 795  
 DB 700 HISPRORPROTPRALAVALAVALAVALAVALAVALAVALAVALAVALAVALAVALA 718  
 QY 746 IGGTCAATATGAGCAATCAICAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 846  
 DB 719 TTPRALAARQPHGLY-----GLUPROVALAVALASERTRIOYLYAIPROVALASPR 745  
 QY 847 -----CAICTGAGCCCTTTCTCTGCT----- 867  
 DB 746 LEUALAARQPHGLYPRORARPROVALARGSERSETARGARGLYALALALALALALAL 755  
 QY 868 GCGAATCTCTGCT 927  
 DB 756 GLYPHPRGGLYALASERPROVALASERTRIOYLYAVALAVALAVALAVALAVALAVAL 775  
 QY 928 -----GCCCCGCGC 946  
 DB 776 PRORARPROVALASERSETPRORARPROVALASERSETPRORARPROVALASERSET 795  
 QY 937 ACTGAGCCCTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 961  
 DB 796 SERPRORARPROVALASERSETPRORARPROVALASERSETPRORARPROVALASER 815  
 QY 992 -CGAGCCGCGCCCT 1049  
 DB 815 GARGPRORARPROVALASERSETPRORARPROVALASERSETPRORARPROVALASER 833  
 QY 1050 -----AGCTGAG 1092  
 DB 833 EUGLYPRORARPROVALASERSETPRORARPROVALASERSETPRORARPROVALASER 852  
 QY 1093 GCGCTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1152  
 DB 852 ILYAUCYHISSEPRORARPROVALASERSETPRORARPROVALASERSETPRORARPRO 871  
 QY 1153 CAAGACAG 1212  
 DB 872 -----TTPRALAARQPHGLY----- 875  
 QY 1213 TGATTCAGATACATACAG 1269  
 DB 875 EUSERGLUPROTPRALAVALAVALAVALAVALAVALAVALAVALAVALAVALAVALA 894  
 QY 1276 GCGAATCTCTGCT 1329  
 DB 895 PRORARPROVALAVALAVALAVALAVALAVALAVALAVALAVALAVALAVALAVALA 900  
 QY 1330 TCAATCT 1389  
 DB 901 -----ARQALAPROLEU----- 904  
 QY 1390 AITCTGCGCCCT 1440  
 DB 905 -----GLIHISARGGLYASERTRIOYLYAIPROVALASERSETPRORARPRO 920  
 QY 1441 CAGCCGCT 1449  
 DB 921 THRVALPR 923  
 RESULT 15  
 SHK1\_HUMAN STANDARD: PRT: 2161 AA.  
 AC Q9Y566; Q9Y566;  
 DT 15-JUN-2002 (Ref: 41, last sequence update)  
 DT 15-JUN-2002 (Ref: 41, last annotation update)  
 DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1)  
 DE (Somaostatin receptor interacting protein) (STRIP interacting protein)  
 DE (SSRIP).  
 OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo  
UX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1: 2 AND 3), AND INTERACTION WITH SSTR2.  
RC TISSUE=Fetal brain, Hippocampus, and Thalamus;  
PX MEDLINE=20020275; PubMed 10551867;  
RA Ziletti H., Hoeberl H.-H., Baeckner D., Richter D., Kretschkamp H.-J.,  
RT "Somatostatin receptor interacting protein defines a novel family of  
PL 1. Biol. Chem. 274:32997-33001(1999)  
RN [2]  
RP REVIEW.  
RX PubMed 10806096;  
RA Sheng M., Kim E.;  
RT "The shank family of scaffold proteins.\*";  
PL 1. Cell Sci. 113:1851-1856(2000)  
CC -1- FUNCTION. Seems to be an adapter protein in the postsynaptic  
density (PSD) of excitatory synapses that interconnects receptors  
of the postsynaptic membrane including NMDA-type and metabotropic  
glutamate receptors via complexes with GRAP/PSD-95 and Homer,  
CC respectively, and the actin based cytoskeleton. May play a role in  
CC the structural and functional organization of the dendritic spine  
CC and synaptic junction.  
CC -1- SUPPLEMENT. May homodimerize via its SAM domain (by similarity).  
CC Interacts with SSTR2 C-terminus via the PDZ domain. Interacts with  
CC SPTAN1, Homer-1 and DISC1/CRAP isoforms 1 and 2 (by similarity).  
CC Is part of a complex with DISC1/PSD-95 and DISC1/CRAP (by  
CC similarity).  
CC -1- SUBCELLULAR LOCATION. Cytoplasmic, postsynaptic density of  
CC neuronal cells (by similarity).  
CC -1- ALTERNATIVE PROTING. 4 isoforms, 1/a (shown here), 2/b and 3, are  
CC produced by alternative splicing.  
CC -1- TISSUE SPECIFICITY. Expressed in brain particularly in the  
CC amygdala, hippocampus, substantia nigra and thalamus. Isoform 2  
CC seems to be expressed ubiquitously.  
CC -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.  
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.  
CC -1- SIMILARITY: CONTAINS 1 PDZ/CRP DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
-----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).  
CC  
CC EMBL: AF163302; ADD45121.1; -  
DR EMBL: AF226728; AAF35887.1; -  
DR DDBP: P06241; 1SHF.  
DR MIM: 604999; -  
DR InterPro: IPR002110; ANK.  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001660; SAM.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF000023; ank; 7.  
DR Pfam: PF000565; pdz; 1.  
DR Pfam: PF00536; SAM; 1.  
DR Pfam: PF000018; SH3; 1.  
DR ProDom: PD000066; SH3; 1.  
DR SMART: SM00248; ANK; 3.  
DR SMART: SM00228; PDZ; 1.  
DR SMART: SM00454; SAM; 1.  
DR SMART: SM00326; SH3; 1.  
DR PROSITE: PS30088; ANK\_REPEAT; 3.  
DR PROSITE: PS30297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE: PS30105; PDZ; 1.  
DR PROSITE: PS30105; SAM\_DOMAIN; 1.  
DR PROSITE: PS50002; SH3; 1.  
DR ANK repeat, SH3 domain, Repeat, Alternative splicing.

[illegible]





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: November 24, 2002, 02:34:10 ; Search time 4797 Seconds

(without alignments)  
11593 801 Million cell updates/sec

Title: US-09-651-150b-1

Perfect score: 1911

Sequence: 1 aaaggaagaaagacagatgac... ttaactgacacacacacit 1911

Scoring table: IDENTITY\_NTC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1455142878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:  
1: gb\_pa\*  
2: gb\_hlg\*  
3: gb\_in\*  
4: gb\_ov\*  
5: gb\_ov\*  
6: gb\_ov\*  
7: gb\_ov\*  
8: gb\_ov\*  
9: gb\_ov\*  
10: gb\_ov\*  
11: gb\_ov\*  
12: gb\_ov\*  
13: gb\_ov\*  
14: gb\_ov\*  
15: gb\_ov\*  
16: gb\_ov\*  
17: gb\_ov\*  
18: gb\_ov\*  
19: gb\_ov\*  
20: gb\_ov\*  
21: gb\_ov\*  
22: gb\_ov\*  
23: gb\_ov\*  
24: gb\_ov\*  
25: gb\_ov\*  
26: gb\_ov\*  
27: gb\_ov\*  
28: gb\_ov\*  
29: gb\_ov\*  
30: gb\_ov\*  
31: gb\_ov\*  
32: gb\_ov\*  
33: gb\_ov\*  
34: gb\_ov\*  
35: gb\_ov\*  
36: gb\_ov\*  
37: gb\_ov\*  
38: gb\_ov\*  
39: gb\_ov\*  
40: gb\_ov\*  
41: gb\_ov\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARY

Result No.	Score	Query Match	Length	DB ID	Description
1	184.4	96.4	2040	AR110400	AR110400 Sequence
2	184.4	96.4	2040	AC006401	AC006401 Homo sapi
3	182.8	95.5	1977	AF057557	AF057557 Homo sapi
4	131.2	68.7	1339	AF062931	AF062931 Homo sapi
5	74.5	39.0	11114	AC068122	AC068122 Homo sapi
6	74.5	39.0	180048	AC098935	AC098935 Homo sapi
7	74.5	39.0	192565	AC023534	AC023534 Homo sapi
8	74.5	39.0	227632	AC023534	AC023534 Homo sapi
9	67.6	35.4	1047	E33201	E33201 P16RL-1 Del
10	67.6	35.4	1047	E33201	E33201 P16RL-1 Del
11	36.4	17.7	160048	AC068122	AC068122 Homo sapi
12	160.6	8.4	58628	AC027719	AC027719 Homo sapi
13	106	5.5	229402	AC023534	AC023534 Homo sapi
14	103.4	5.4	61451	AC101514	AC101514 Mus muscu
15	100.4	5.3	104087	AC125973	AC125973 Rattus no
16	89.4	3.6	61451	AC101514	AC101514 Mus muscu
17	62.8	3.3	125020	AF429315	AF429315 Homo sapi
18	60.4	3.2	125020	AF429315	AF429315 Homo sapi
19	59.6	3.1	104087	AC125973	AC125973 Rattus no
20	56.2	2.9	7218	16444	16444 Sequence 14
21	52.6	2.8	156093	AC128528	AC128528 Rattus no
22	52.2	2.7	1755	AB048834	AB048834 Mus muscu
23	51.2	2.7	115666	AC105744	AC105744 Oryza sat
24	51.4	2.7	115666	AC105744	AC105744 Oryza sat
25	50	2.6	37305	HS344F5	HS344F5 Mus muscu
26	50	2.6	195880	AC128498	AC128498 Homo sapi
27	49.6	2.6	43147	SC4A10	SC4A10 Homo sapi
28	49.6	2.6	144552	AC112723	AC112723 Homo sapi
29	49.2	2.6	57975	AC128498	AC128498 Homo sapi
30	49.2	2.6	61958	AC121467	AC121467 Rattus no
31	48.8	2.6	55591	AC087813	AC087813 Homo sapi
32	48.6	2.5	44764	AC037442	AC037442 Homo sapi
33	48.6	2.5	132180	AC125927	AC125927 Rattus no
34	48.6	2.5	136828	AC127769	AC127769 Rattus no
35	48.4	2.5	16376	AC112902	AC112902 Rattus no
36	48.2	2.5	2430	AR204703	AR204703 Sequence
37	48.2	2.5	2904	AB005293	AB005293 Homo sapi
38	48.2	2.5	2979	HC031084	HC031084 Homo sapi
39	48.2	2.5	135426	AC120061	AC120061 Rattus no
40	48.2	2.5	168312	AC128552	AC128552 Rattus no
41	47.8	2.5	173278	AC129135	AC129135 Rattus no
42	47.8	2.5	178428	AC108601	AC108601 Rattus no
43	47.6	2.5	14340	HS222E13	HS222E13 Human DNA S
44	47.6	2.5	160725	HS919B11	HS919B11 Homo sapi
45	47.2	2.5	198273	AC114897	AC114897 Mus muscu

## ALIGNMENTS

RESULT 1  
LOCUS AR110400  
DEFINITION Sequence 1 from patent US 6114515.  
ACCESSION AR110400  
VERSION AR110400.1 GI:12826676  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2040)  
AUTHORS Wu, S., Sweet, R. W. and Trunch, A.  
TITLE FIG. 1, a member of immunoglobulin gene superfamily  
FEATURES  
Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a





ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Cladista; Vertebrata; Eutelestomi,  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 2040)  
 TITLE Syjjan, O., Raymond, W. S. and Allemsaqet, T.  
 PIGRI-1 belonging to the immunoglobulin gene superfamily  
 JOURNAL Patent: JP 199151094-A 1 08-JUN-1999;  
 SMITHKLINE BEECHAM CORP  
 OS Homo sapiens (human)  
 PN JP 199151094-A/1  
 PD 08-JUN-1999  
 PE 25-AUG-1998 JP 1998339287  
 PR 30-OCT-1997 US 08/961564, 45-AUG-1997 US 60/056935 P1  
 PU SUJIAN O, RAYMOND W, SWEET, ALFENSCHEID TRUNEH  
 PC C12N15/09, A61K39/00, A61K39/395, A61K45/00, A61K48/00,  
 PC C07K14/47,  
 PC C07K16/18, C12N5/10, C12P21/02, C12N15/09, C12R1:91, C12N5/10,  
 PC C12R1:91),  
 PC (C12P21/02, C12R1:91), C12N15/00, A61K37/02, C12N5/00, C12N15/00,  
 PC C12R1:91),  
 PC (C12N5/00, C12R1:91)  
 PG  
 CC  
 FI  
 FT  
 FT  
 KEY Location/Qualifiers  
 source 1..2040  
 location/Qualifiers  
 1..2040  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

Query Match 96.4%; Score 1842.4; DB 6; Length 2040;  
 Best Local Similarity 99.2%, Pred. No. 0.  
 Matches 1894; Conservative 0; Mismatches 11; Indels 5; Gaps 4;

Y 5 CACIACACGCGTGTCTCCATCCCTCTCAAGGCGCTCTGGATCCACCTTGCACCTCA 64  
 Db 77 GATTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 136  
 Y 65 GAATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 124  
 Db 137 GAAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 196  
 Y 125 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 184  
 Db 197 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 256  
 Y 185 CCACCTTCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 244  
 Db 257 CCATTTCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 316  
 Y 245 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 304  
 Db 317 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 376  
 Y 305 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 364  
 Db 377 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 436  
 Y 365 GACAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 424  
 Db 437 GACAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 496  
 Y 425 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 484  
 Db 497 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 556  
 Y 485 AGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 544  
 Db 557 AGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 616  
 Y 545 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 604

Y 617 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 676  
 Y 605 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 664  
 Db 677 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 736  
 Y 665 GGTACAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 724  
 Db 737 GGTACAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 796  
 Y 725 GGTACAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 784  
 Db 797 GGTACAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 856  
 Y 785 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 844  
 Db 857 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 916  
 Y 845 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 904  
 Db 917 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 976  
 Y 905 AGGAGGAAAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 964  
 Db 977 AGGAGGAAAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1036  
 Y 965 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1024  
 Db 1037 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1096  
 Y 1025 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1084  
 Db 1097 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1156  
 Y 1085 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1144  
 Db 1157 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1216  
 Y 1145 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1204  
 Db 1217 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1276  
 Y 1205 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1264  
 Db 1277 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1336  
 Y 1265 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1324  
 Db 1337 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1396  
 Y 1325 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1384  
 Db 1397 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1456  
 Y 1385 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1444  
 Db 1457 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1516  
 Y 1445 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1504  
 Db 1516 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1576  
 Y 1505 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1564  
 Db 1576 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1636  
 Y 1565 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1624  
 Db 1636 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1696  
 Y 1625 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1684



845 ACCATCTGGGCTTTTCTGCTGAGATCTGGAGTGGATGTTGAAAAGAGGCTTGA 904  
 925 ACCATCTGGGCTTTTCTGCTGAGATCTGGAGTGGATGTTGAAAAGAGGCTTGA 964  
 905 AG 964  
 985 AG 1044  
 965 AGCTTCTAG 1024  
 1045 AGCTTCTAG 1104  
 1025 AGCTTCTAG 1084  
 1105 AGCTTCTAG 1164  
 1085 AGCTTCTAG 1144  
 1165 AGCTTCTAG 1224  
 1145 AGCTTCTAG 1204  
 1225 AGCTTCTAG 1284  
 1205 AGCTTCTAG 1264  
 1285 AGCTTCTAG 1344  
 1265 AGCTTCTAG 1324  
 1345 AGCTTCTAG 1404  
 1325 AGCTTCTAG 1384  
 1405 AGCTTCTAG 1464  
 1385 AGCTTCTAG 1444  
 1465 AGCTTCTAG 1524  
 1445 AGCTTCTAG 1504  
 1524 AGCTTCTAG 1584  
 1505 AGCTTCTAG 1564  
 1584 AGCTTCTAG 1644  
 1565 AGCTTCTAG 1624  
 1644 AGCTTCTAG 1702  
 1625 AGCTTCTAG 1684  
 1703 AGCTTCTAG 1762  
 1685 AGCTTCTAG 1742  
 1763 AGCTTCTAG 1822  
 1803 AGCTTCTAG 1862  
 1883 AGCTTCTAG 1942  
 1863 AGCTTCTAG 1885  
 1943 AGCTTCTAG 1965

RESULT 4  
 AF057557 1339 bp mRNA linear pri 04-JUN-1998  
 LOCUS Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.  
 DEFINITION  
 AF057557  
 VERSION  
 AF057557.1 GI:3169292  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens.  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE  
 1 (bases 1 to 1339)  
 Hinooshi, Y., Lorenz, J., Kitada, S. I., Fisher, J., Labarge, M.,  
 King, H. Z., Franke, U., Reed, J. C., Kinoshita, S. and Nolan, G. P.  
 TOSO, a cell surface, specific regulator of Fas-induced apoptosis  
 in T cells  
 Immunology 94 (4), 461-471 (1998)  
 MEDLINE  
 PUBMED  
 98246048  
 9586536  
 REFERENCE  
 2 (bases 1 to 1339)  
 Labarge, M. and Hinooshi, Y.  
 Direct Submission  
 Submitted (01-APR-1999) Molecular Pharmacology, Stanford  
 University, 300 Pasteur Dr., Palo Alto, CA 94304, USA  
 FEATURES  
 source  
 1..1339  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1q31-q32"  
 /cell\_line="Jurkat"  
 /cell\_type="T cell"  
 /note="TOSO.2 cDNA, cloned by PCR using primers designed  
 from the original TOSO.1 cDNA sequence"  
 1..1339  
 /gene="TOSO"  
 20..1192  
 /gene="TOSO"  
 /note="transmembrane protein; Ig superfamily member"  
 /codon\_start=1  
 /product="anti-Fas-induced apoptosis"  
 /protein\_id="AAC18830.1"  
 /db\_xref="GI:3169293"  
 /translation="MDPMLPFLVPSGALRIIPVKVKGHIGGSVILKPIIPKMH  
 KIVLKEHMASSGIGIVSVLINFLEKAYKQKVLKQVKKNLIVLVLIHLSUSQV  
 AGCGMNTDKRKTOKVTLNHSSEYDSMEHPPIPKMFLPIPLQMAVASSKRV  
 LPVITPAPQKRVPPVHNSPTQTITPVPVSSVAGKPPFLPTTASVSALEGL  
 LKRPQTSYNNHILRLHQRALDYGSGSGREGGFHLLPILLLPLALLGLGVKAV  
 ERKALSPFAVFLAVPMRALESSKPKGSPKPSQNNIVSAVPRKRGDAGTGAP  
 VPQCAPLPAPAPQVSESPWIMHSLKISCEYSLVHQAAMHEDSDIDYINVA"  
 127  
 /gene="TOSO"  
 /note="occurs in TOSO.1 cDNA"  
 /replac="a"  
 1015  
 /gene="TOSO"  
 /note="occurs in TOSO.1 cDNA"  
 /replac="g"  
 variation  
 BASE COUNT 309 a 432 c 340 g 258 t  
 ORIGIN  
 Query Match 68.7%; Score 1312.2; DH 9; length 1339;  
 Host local similarity 99.6%; Prod. No. 0;  
 Matches 1336; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
 QY 55 TTTAAATTAAG 114  
 DB 1 TTTAAATTAAG 60  
 QY 115 AGCGGCTTGGAG 174  
 DB 61 AGCGGCTTGGAG 120

QY 175 GATCAATGGGACCTTCTTGAATGATGAGAGATATATCTGGGGGAGATGGCGG 234  
 DB 121 GATCAATGGGACCTTCTTGAATGATGAGAGATATATCTGGGGGAGATGGCGG 180  
 QY 235 ATCTGCAAGATGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 294  
 DB 181 ATCTGCAAGATGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 240  
 QY 295 GCGAGTACTCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 354  
 DB 241 GCGAGTACTCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 300  
 QY 355 GAGCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 414  
 DB 301 GAGCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 460  
 QY 415 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 474  
 DB 361 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 420  
 QY 475 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 534  
 DB 421 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 480  
 QY 535 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 594  
 DB 481 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 540  
 QY 595 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 654  
 DB 541 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 600  
 QY 655 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 714  
 DB 601 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 660  
 QY 715 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 774  
 DB 661 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 720  
 QY 775 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 834  
 DB 721 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 780  
 QY 835 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 894  
 DB 781 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 840  
 QY 895 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 954  
 DB 841 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 900  
 QY 955 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1014  
 DB 901 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 960  
 QY 1015 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1074  
 DB 961 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1020  
 QY 1075 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1134  
 DB 1021 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1080  
 QY 1135 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1194  
 DB 1081 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1140  
 QY 1195 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1254  
 DB 1141 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1200

QY 1255 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1314  
 DB 1201 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1260  
 QY 1315 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1374  
 DB 1261 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1319  
 QY 1375 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1439  
 DB 1320 GAGTCAAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAG 1280

RESULT 5  
 AY062931/G  
 DEFINITION Homo sapiens interleukin 24 (IL24) gene, complete cds.  
 ACCESSION AY062931  
 VERSION AY062931.1 GI:17104816  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1114)  
 RIDER, M.D., ARMEL, F.Z., CARLINGTON, D.P., CHUNG, M. W., LEE, K. L.,  
 POEL, C. L., TOBE, E. J., YI, V., and Nickerson, D. A.  
 Direct Submission  
 Submitted (13-NOV-2001) Molecular Biotechnology, University of  
 Washington, 1705 NE Pacific, Seattle, WA 98195, USA  
 To cite this work please use: SeattleSNPs, NHLBI Program for  
 Genomic Applications, UW-PHCRC, Seattle, WA (URL:  
 http://pga.mbl.washington.edu).  
 Location/Qualifiers  
 1..1114  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 543..856  
 /rpl\_family="MER1\_type"  
 /rpl\_type-dispersed  
 627  
 /frequency="0.01"  
 /replace="a"  
 634  
 /frequency="0.10"  
 /replace="c"  
 806  
 /frequency="0.05"  
 /replace="g"  
 884..925  
 /rpl\_family="MER1\_type"  
 /rpl\_type-dispersed  
 1619..1620  
 /frequency="0.07"  
 /replace="a"  
 1625..1759  
 /rpl\_family="SINE"  
 /rpl\_type-dispersed  
 1627  
 /frequency="0.02"  
 /replace="c"  
 1889  
 /frequency="0.06"  
 /replace="g"  
 1897..8320  
 /gene="IL24"  
 /note(1897..2068,2184..2359,3775..3970,4751..4813,  
 5950..6108,6454..6528,7412..8320)  
 /gene="IL24"  
 /product="Interleukin 24"  
 2049  
 /gene="IL24"  
 /frequency="0.06"

CDS  
 /replace="t"  
 join(2286..2329,3775..3970,4751..4813,5950..6104,  
 6454..6528,7432..7515)  
 /gene="IL24"  
 /codon\_start=1  
 /product="interleukin 24"  
 /protein\_id="AAU34146.1"  
 /db\_xref="GI:17104817"  
 /translation="MNFQORLOSITLAPFCEPLLATASQMQVYVPLGFTLLIKS  
 QVSAAGVLEHSPVQKSVQKLTWATWAKVTMCAQNTISAKLLQCELVQVSD  
 AESCYLVHLEFLYKIVEKHNRTVEVRLKSFSTLANNFVLIVSLQPSQENEMF  
 SIKDSAHRRPLIFRRAPKQIDVRAALTKALGEVDILLIMOKFYKI"  
 2324..2640  
 /rpt\_family="SINE"  
 /rpt\_type-dispersed  
 2461  
 variation  
 /gene="IL24"  
 /frequency="0.90"  
 /replace="a"  
 2546  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="a"  
 2625..3110  
 /gene="IL24"  
 /note="Region not scanned for variation"  
 3221  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="c"  
 3260  
 variation  
 /gene="IL24"  
 /frequency="0.37"  
 /replace="t"  
 3292  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="a"  
 3352  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="t"  
 3466  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="a"  
 3517  
 variation  
 /gene="IL24"  
 /frequency="0.02"  
 /replace="a"  
 3530  
 variation  
 /gene="IL24"  
 /frequency="0.05"  
 /replace="a"  
 3551  
 variation  
 /gene="IL24"  
 /frequency="0.99"  
 /replace="a"  
 3625  
 variation  
 /gene="IL24"  
 /frequency="0.37"  
 /replace="c"  
 4061  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="a"  
 4069  
 variation  
 /gene="IL24"  
 /frequency="0.36"  
 /replace="t"  
 4082  
 variation  
 /gene="IL24"  
 /frequency="0.02"

repeat\_region  
 /replace="c"  
 4094..4167  
 /rpt\_family="SINE"  
 /rpt\_type-dispersed  
 4243  
 variation  
 /gene="IL24"  
 /frequency="0.37"  
 /replace="a"  
 4254  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="g"  
 4332  
 variation  
 /gene="IL24"  
 /frequency="0.64"  
 /replace="a"  
 4585  
 variation  
 /gene="IL24"  
 /frequency="0.17"  
 /replace="a"  
 4602  
 variation  
 /gene="IL24"  
 /frequency="0.38"  
 /replace="t"  
 4943  
 variation  
 /gene="IL24"  
 /frequency="0.02"  
 /replace="a"  
 5020  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="a"  
 5388  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="a"  
 5392  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="t"  
 5519  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="c"  
 5637..5821  
 repeat\_region  
 /rpt\_family="SINE"  
 /rpt\_type-dispersed  
 6016  
 variation  
 /gene="IL24"  
 /frequency="0.35"  
 /replace="c"  
 6020  
 variation  
 /gene="IL24"  
 /frequency="0.02"  
 /replace="g"  
 6037  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="c"  
 6115..6280  
 repeat\_region  
 /rpt\_family="SINE"  
 /rpt\_type-dispersed  
 6240  
 variation  
 /gene="IL24"  
 /frequency="0.01"  
 /replace="a"  
 6282  
 variation  
 /gene="IL24"  
 /frequency="0.37"

Query Match 39.0% Score 745; DB 9; Length 11114;  
 Best Local Similarity 98.2% Pred. No. 1,36-177;  
 Matches 785; Conservativ. 0; Mismatches 10; Indels 4; Gaps 3;





Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

#### Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below.

The electronically-digested sequence consists of both insert and vector. In order to accurately represent the entire circular BAC, small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BgIII		EcoRI		HindIII	
SeqDerMap	Fingerprint	SeqDerMap	Fingerprint	SeqDerMap	Fingerprint
3569	3709	8696	8616	3665	3596
2067	4000	6	<800	6382	6620
5849	5832	102	<800	512	<800
9284	9171	1103	1094	449	<800
821	792	961	951	4965	4877
770	<800	1103	1094	814	776
2354	2403	80	<800	23	<800
3661	3867	2275	2290	5727	5699
183	<800	1782	1810	2120	2127
975	1001	13922	13969	900	932
310	<800	1449	1431	3852	3871
1379	1347	2677	2695	1097	1136
641	<800	991	951	1530	1512
2840	2860	464	<800	1962	1952
3227	3296	1200	1209	366	<800
724	<800	4852	4871	3148	3167
3986	4005	36	<800	2430	2403
3092	3080	924	951	3040	3022
1248	1263	5333	5392	13940	13551
5297	5194	9804	9608	31	<800
630	<800	1526	1529	6384	6620

14593	15092	341	<800	1749	1747
3023	3080	7007	7047	1184	1136
6232	6074	3763	3727	3579	3566
2107	2060	8262	8127	173	<800
285	<800	5122	5033	2341	2403
2406	2403	253	<800	6231	6239
4382	4328	3715	3727	963	932
7405	7374	5599	5634	38	<800
1488	1508	2268	2290	234	<800
957	884	8470	8616	497	<800
671	855	6087	6102	3196	3167
2278	2403	8573	8616	5384	5304
6090	6074	629	<800	1584	1512
296	<800	12	<800	20587	20250
2730	2735	1634	1665	19617	20250
1542	1508	7111	7037	2739	2744
643	<800	1825	1810	174	<800
1626	1674	3347	3347	3542	3596
3469	3503	5424	5392	6530	6620
2539	2403	53	<800	172	<800
3827	3867	11030	10924	57	<800
1976	2000	2861	2872	760	776
5618	5486	4572	4534	7921	7962
3904	3867	1413	1431	4835	4877
1248	1263	2444	2432	743	776
3278	3296	2401	2432	743	776
1284	1263	325	<800	1289	1306
510	<800	178	<800	5345	5304
9617	9614	322	<800	1619	1615
1170	1153	500	<800	373	<800
10558	10465	4358	4271	23	<800
1695	1674	719	<800	1357	1306
5976	5832	8202	8127	6531	6620
2067	2060	25	<800	3919	3871
5040	4859	4091	4050	2835	2836
991	1001	1208	1209	6702	6620





*	7774	9425:	contig of 1652 bp	in	length
*	9426	9525:	gap of 103 bp	in	length
*	9526	10997:	contig of 1472 bp	in	length
*	10944	11047:	gap of 100 bp		
*	10944	12740:	contig of 1543 bp	in	length
*	12741	12840:	gap of 100 bp		
*	12841	14467:	contig of 1627 bp	in	length
*	14468	14567:	gap of 100 bp		
*	14568	16980:	contig of 2413 bp	in	length
*	16981	17080:	gap of 100 bp		
*	17081	19176:	contig of 2096 bp	in	length
*	19177	19276:	gap of 100 bp		
*	19277	21882:	contig of 2606 bp	in	length
*	21881	21982:	gap of 100 bp		
*	21981	23826:	contig of 1844 bp	in	length
*	23827	23926:	gap of 100 bp		
*	23927	25561:	contig of 2635 bp	in	length
*	25562	26661:	gap of 100 bp		
*	26662	28846:	contig of 2185 bp	in	length
*	28847	28946:	gap of 100 bp		
*	28947	32495:	contig of 3549 bp	in	length
*	32496	32595:	gap of 100 bp		
*	32596	34950:	contig of 2355 bp	in	length
*	34951	35050:	gap of 100 bp		
*	35051	38193:	contig of 3143 bp	in	length
*	38194	38293:	gap of 100 bp		
*	38294	41862:	contig of 3569 bp	in	length
*	41863	41962:	gap of 100 bp		
*	41963	45898:	contig of 3836 bp	in	length
*	45899	45998:	gap of 100 bp		
*	45999	50904:	contig of 4405 bp	in	length
*	50905	51004:	gap of 100 bp		
*	51005	55346:	contig of 4442 bp	in	length
*	55347	55446:	gap of 100 bp		
*	55447	61336:	contig of 5940 bp	in	length
*	61337	61436:	gap of 100 bp		
*	61437	65808:	contig of 4370 bp	in	length
*	65809	65906:	gap of 100 bp		
*	65907	71648:	contig of 5742 bp	in	length
*	71649	71748:	gap of 100 bp		
*	71749	74886:	contig of 3138 bp	in	length
*	74887	74986:	gap of 100 bp		
*	74987	79878:	contig of 4892 bp	in	length
*	79879	79978:	gap of 100 bp		
*	79979	864280:	contig of 4302 bp	in	length
*	84281	84380:	gap of 100 bp		
*	84381	89873:	contig of 5493 bp	in	length
*	89874	89973:	gap of 100 bp		
*	89974	97168:	contig of 7195 bp	in	length
*	97169	97268:	gap of 100 bp		
*	97269	101906:	contig of 4638 bp	in	length
*	101907	102906:	gap of 100 bp		
*	102907	109439:	contig of 7433 bp	in	length
*	109440	109539:	gap of 100 bp		
*	109540	117592:	contig of 8053 bp	in	length
*	117593	117692:	gap of 100 bp		
*	117693	126037:	contig of 8345 bp	in	length
*	126038	126137:	gap of 100 bp		
*	126138	135140:	contig of 9003 bp	in	length
*	135141	135240:	gap of 100 bp		
*	135241	146384:	contig of 11144 bp	in	length
*	146385	146484:	gap of 100 bp		
*	146485	160352:	contig of 13868 bp	in	length
*	160353	160452:	gap of 100 bp		
*	160453	173458:	contig of 13006 bp	in	length
*	173459	173558:	gap of 100 bp		
*	173559	191605:	contig of 18047 bp	in	length
*	191606	191705:	gap of 100 bp		
*	191706	223304:	contig of 37597 bp	in	length
Location/Qualifiers					
1 238302					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
FEATURES					
SOURCE					

[illegible]





```

* 32792 32891: gap of unknown length
* 32892 38895: contig of 6004 bp in length
* 38896 38996: gap of unknown length
* 38996 44306: contig of 5311 bp in length
* 44307 44406: gap of unknown length
* 44407 51589: contig of 7183 bp in length
* 51590 51689: gap of unknown length
* 51690 58591: contig of 6902 bp in length
* 58592 58692: gap of unknown length
* 58692 63442: contig of 4751 bp in length
* 63443 63542: gap of unknown length
* 63543 70554: contig of 7012 bp in length
* 70555 70654: gap of unknown length
* 70655 77794: contig of 7140 bp in length
* 77795 77894: gap of unknown length
* 77895 85842: contig of 7947 bp in length
* 85842 85942: gap of unknown length
* 85942 93327: contig of 7286 bp in length
* 93328 93327: gap of unknown length
* 93328 100871: contig of 7544 bp in length
* 100872 100971: gap of unknown length
* 100972 112141: contig of 11170 bp in length
* 112142 112341: gap of unknown length
* 112342 126358: contig of 14117 bp in length
* 126359 126458: gap of unknown length
* 126459 141767: contig of 15309 bp in length
* 141768 141867: gap of unknown length
* 141868 155927: contig of 14060 bp in length
* 155928 156027: gap of unknown length
* 156028 180048: contig of 24021 bp in length.

FEATURES
Source
1 180048
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="Pp11-258p22"

BASE COUNT 49152 a 39956 c 40116 g 48114 t 2710 others

ORIGIN
Query Match 17.7% Score 338.6, DB 2, length 180048,
Host local similarity 98.8% Prod No. 1 5e-74:
Matches 341: Conservative 0; Mismatches 4; Indels 0; Gaps 0:

UY 102 TCTCTTATGATAGAGGTTTATGATTTTAAATTAAGATAGAGATGATG 161
DB 89183 TCCCTGAGATGACGGCCCTGACGATCCTCCAGAGTAAGGTAGAGCGGACCTCG 89242
UY 162 GCGATGAGTATTAATATGATTTTATGATTTTATGATTTTATGATTTTATG 221
DB 89243 GCGATGAGTATTAATATGATTTTATGATTTTATGATTTTATGATTTTATG 89302
UY 222 GCGATGAGTATTAATATGATTTTATGATTTTATGATTTTATGATTTTATG 281
DB 89303 GCGATGAGTATTAATATGATTTTATGATTTTATGATTTTATGATTTTATG 89362
UY 282 CAGAAATGAGGAGGAGTATGATTAATAGGATTAAGGATTTGTTAGTGG 341
DB 89363 CAGAAATGAGGAGGAGTATGATTAATAGGATTAAGGATTTGTTAGTGG 89422
UY 342 AGTAAATGATTAATAGGATTTGATGATTTGATGATTTGATGATTTGATG 401
DB 89423 AGTAAATGATTAATAGGATTTGATGATTTGATGATTTGATGATTTGATG 89482
UY 402 CAGAAATGAGGAGGAGTATGATTAATAGGATTTGATGATTTGATGATTTGATG 446
DB 89483 CAGAAATGAGGAGGAGTATGATTAATAGGATTTGATGATTTGATGATTTGATG 89527

RESULT 12
AC027719 58628 bp DNA linear HTG 01-AIR 2003
DEFINITION Homo sapiens chromosome 8 clone Pp11-258p22 map 8, LOW PASS SEQUENCE
SAMPLING.

```

```

ACCESSION AC027719
VERSION AC027719.1 GI:7382531
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 58628)
AUTHORS Birren, B., Linton, J., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 8, clone Pp11-258p22
REFERENCE 2 (bases 1 to 58628)
AUTHORS Birren, B., Linton, J., Nusbaum, C., Lander, E., Abramson, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F., Boguski, M., Bouckge, K., Brown, A., Burdett, G., Campilioni, A., Cooke, P., DeArnell, K., Dwyer, K., Dwyer, J., Dwyer, S., Dwyer, M., Dwyer, M., Felleman, P., Fitzhugh, W., Gade, D., Gaidanov, J., Gaidanov, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heath, A., Horton, L., Howland, J. C., Iliu, I., Johnson, K., Jones, C., Kann, L., Karas, A., Klein, J., Lacombe, K., Lamazares, P., Landers, T., Lebeck, J., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margolis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R., Melidim, J., Menus, L., Mihova, T., Miranda, C., Miura, Y., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisanic, C., Pollara, V., Raymond, C., Riley, P., Roy, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tessile, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.
DIRECT SUBMISSION
COMMENT
JOURNAL
TITLE
JOURNAL
COMMENT
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome Research, 220 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project information
Center project name: 17449
Center clone name: 258_p_9
* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
672 771: gap of 100 bp
772 1469: contig of 698 bp in length
1470 1569: gap of 100 bp
1570 2461: contig of 692 bp in length
2462 2361: gap of 100 bp
2362 3051: contig of 666 bp in length
3052 3151: gap of 100 bp
3152 3816: contig of 665 bp in length
3817 3916: gap of 100 bp
3917 4597: contig of 681 bp in length
4598 4697: gap of 100 bp
4698 5395: contig of 698 bp in length
5396 5495: gap of 100 bp

```











LOCUS AC125973 104087 bp. FNA Linear HTG 1st-2002  
 DEFINITION Rattus norvegicus clone CH230-64N5. \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 AC125973  
 VERSION AC125973.1 GI:21671794  
 KEYWORDS HTG, HTGS, PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Rattus.

REFERENCE  
 1 (bases 1 to 104087)  
 Muzny, D.M., Adams, R., Adinolfi, A., Aliosman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.P., Ayala, M., Banks, T., Barbata, J., Benton, J., Blomquist, K., Blankenship, K., Bonin, D., Bouck, J., Brown, S., Brown, M., Brown, E., Brown, M., Bryant, N.P., Hubay, C., Hurch, P., Hurkett, C., Hurkett, K.I., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, P., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.P., Cox, C., Coyle, M.D., Dalborge, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dedrich, D.A., Delaney, K.R., Delgado, G., Dem, A.I., Ding, Y., Ditt, H., Douthett, K.J., Draper, H., Dugan-Rocha, S., Durkin, K.I., Earnhart, C., Edgar, B., Edwards, C.C., Elhaj, C., Escott, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P., Gabel, J., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, P., Gorielli, J., Goveara, W., Goumar, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., Hernandez, J., Hernandez, O., Hodges, A., Hughes, M., Hollway, C., Hollins, B., Hombl, F., Howard, S., Huber, J., Bullyk, S., Hume, J., Jackson, L.E., Jarboe, R., Jia, Y., Johnson, R., Jolliffe, S., Joudah, S., Karlsson, E., Kelly, S., Khan, N., King, I., Korvah, T., Kovar, C., Kravtsov, D., Kureshi, A., Landry, N., Leal, R., Lewis, J.C., Lewis, J., Li, J., Li, Z., Litchfield, O., Liu, C., Liu, J., Liu, W., Lonsdale, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, P., Martindale, A., Martinez, E., Massey, E., Mathew, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Monahan, K., Morgan, M., Morris, S., Moser, M., Ned, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, R., Nwokwu, S., Ogih, M., Okumura, G., Oran, N., Oviatt, R., Papp, A., Payton, H., Peck, I., Peters, L., Plickens, R., Plinius, E., Pui, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojehokan, I., Polle, M., Ruiz, S., Savery, S., Scherer, S., Scott, G., Shen, H., Shoshitaishvili, N., Sisson, I., Sodergren, E., Soudike, I., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatoch, A., Tabor, P., Tamorisa, A., Tamorisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umanil, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, P., Wang, C., Wang, S., Ward-Moore, S., Warren, P., Washington, C., Wallington, S., Williams, G., Williamson, A., Wierzyk, P., Worden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Chib, R.

TITLE Direct Submission  
 JOURNAL unpublished  
 REFERENCE 2 (bases 1 to 104087)  
 AUTHOR Worley, K.C.  
 JOURNAL Direct Submission  
 REFERENCE Submitted (02 Jul-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 JOURNAL 3 (bases 1 to 104087)  
 AUTHOR Worley, K.C.  
 JOURNAL Direct Submission  
 REFERENCE Submitted (13 Jul-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 JOURNAL genome center  
 COMMENT Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 Center project name: GINE

Center clone name: CH230-64N5  
 ----- Summary Statistics -----  
 Sequencing vector: Plasmid  
 Chemistry: dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap; version 0.960329  
 Consensus quality: 52404 bases at least Q40  
 Consensus quality: 55275 bases at least Q30  
 Consensus quality: 58433 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 NOTE: This is a "working draft" sequence. It currently consists of 51 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	1001:	contig of 1001 bp in length
1002	1101:	gap of unknown length
1102	2424:	contig of 1323 bp in length
2425	2524:	gap of unknown length
2525	4131:	contig of 1607 bp in length
4132	4231:	gap of unknown length
4232	5397:	contig of 1166 bp in length
5398	5497:	gap of unknown length
5498	6828:	contig of 1331 bp in length
6829	6929:	gap of unknown length
6929	8111:	contig of 1183 bp in length
8112	8211:	gap of unknown length
8212	9058:	contig of 1747 bp in length
9059	10058:	gap of unknown length
10059	11180:	contig of 1122 bp in length
11181	11280:	gap of unknown length
11281	12344:	contig of 1064 bp in length
12345	12444:	gap of unknown length
12445	14126:	contig of 1682 bp in length
14127	14226:	gap of unknown length
14227	15269:	contig of 1043 bp in length
15270	15370:	gap of unknown length
15370	16422:	contig of 1053 bp in length
16423	16522:	gap of unknown length
16523	18116:	contig of 1594 bp in length
18117	18216:	gap of unknown length
18217	19531:	contig of 1315 bp in length
19532	19631:	gap of unknown length
19632	20729:	contig of 1098 bp in length
20730	21829:	gap of unknown length
21830	21860:	contig of 1031 bp in length
21861	21960:	gap of unknown length
21961	23506:	contig of 1600 bp in length
23507	23666:	gap of unknown length
23667	25498:	contig of 1832 bp in length
25499	25598:	gap of unknown length
25599	27057:	contig of 1459 bp in length
27058	27158:	gap of unknown length
27159	28982:	contig of 1825 bp in length
28983	29082:	gap of unknown length
29083	30148:	contig of 1066 bp in length
30149	30248:	gap of unknown length
30249	31700:	contig of 1452 bp in length
31701	31800:	gap of unknown length
31801	33292:	contig of 1492 bp in length
33293	33392:	gap of unknown length
33393	34632:	contig of 1240 bp in length
34633	34732:	gap of unknown length
34733	36196:	contig of 1464 bp in length
36197	36296:	gap of unknown length
36297	37447:	contig of 1051 bp in length
37448	37448:	gap of unknown length
37449	39820:	contig of 2373 bp in length
39821	39920:	gap of unknown length



.

.

.

.

.

.

.

.













P1 Sweet RW, Truneh A, Wu S;  
 XX MPI: 1999-1926677  
 DR P-PSDB: AAY05002.  
 XX  
 PT New polypeptides encoding PIGRL-1 useful for treating diseases such  
 XX as X-linked Severe Combined Immunodeficiency  
 PS  
 XX Disclosure, Page 8, 26pp. English

CC This sequence encodes the human PIGRL-1 protein of the invention  
 CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
 CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined  
 CC Immunodeficiency (XSCID) and Iga deficiency. These diseases can be  
 CC diagnosed or susceptibility to them predicted by: (1) determining whether  
 CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
 CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
 CC Patients deficient in PIGRL-1 can be treated by administering either the  
 CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
 CC Patients with excessive expression or activity of PIGRL-1 can be treated  
 CC by administering an antagonist of PIGRL-1, an antisense nucleic acid  
 CC molecule which inhibits the expression of PIGRL-1 or administering  
 CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
 CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
 CC with a candidate compound in the presence of a signal system and noting  
 CC the candidate as an agonist if a signal is produced. The same method can  
 CC be used to identify antagonists of PIGRL-1 but the presence of an  
 CC antagonist is indicated by a decrease in production of the signal.  
 CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
 CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
 CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
 CC This information may then be correlated with the incidence of autoimmune  
 CC disease in those patients to identify whether the mutation causes the  
 CC disease.

XX  
 XX Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T; 12 other;

Query Match 35.4% Scale 676, DB 20 Length 1047;

Best Local Similarity 94.6% Pred. No. 4,6e-180;

Matches /85: Conservative 0; Mismatches 34; Indels 11; Gaps 8;

QY 5 GACTAGCAGCCGCTGCTCCATCCCCCTGCTGAGGGGCTGTCGATGACCTTCGACTTA 64  
 DB 81 GACTAGCAGCCGCTGCTCCATCCCCCTGCTGAGGGGCTGTCGATGACCTTCGACTTA 140  
 QY 65 GAAGGCAATATGATTTTGGTTTGGTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 124  
 DB 141 GAAGGCAATATGATTTTGGTTTGGTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 200  
 QY 125 AGCATCTTCGCAAGTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 184  
 DB 201 AGCATCTTCGCAAGTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 260  
 QY 185 CCACTTCCTGAAATGATGAGATATATCTGTCGCCGACATCCCTGATCTGCAACA 244  
 DB 261 CCACTTCCTGAAATGATGAGATATATCTGTCGCCGACATCCCTGATCTGCAACA 320  
 QY 245 TGTGTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 304  
 DB 321 TGTGTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 380  
 QY 305 CTGGAAGCAATACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 364  
 DB 381 CTGGAAGCAATACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 440  
 QY 365 GATAGAGGAGATTTATGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424  
 DB 441 GATAGAGGAGATTTATGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 499  
 QY 425 GTATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 484  
 DB 500 GTATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 559

QY 485 ACHCAAAAGGATTCATGATGATGATGATGATGATGATGATGATGATGATGATG 544  
 DB 560 ACHCAAAAGGATTCATGATGATGATGATGATGATGATGATGATGATGATGATG 619  
 QY 545 AATTCGTAAAC--CAAGATTAATGATGATGATGATGATGATGATGATGATGATG 601  
 DB 620 AATTCGTAAACCAAGATTAATGATGATGATGATGATGATGATGATGATGATG 679  
 QY 600 CATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 659  
 DB 680 CATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 739  
 QY 660 TACCAAGTACCAAGCCGCC--AACTTCGATGATGATGATGATGATGATGATGATG 718  
 DB 740 TACCAAGTACCAAGCCGCCGAAATTCGATGATGATGATGATGATGATGATGATG 799  
 QY 719 CTGAAAGGAGATTTATGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 775  
 DB 800 CTGAAAGGAGATTTATGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 858  
 QY 776 AGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 825  
 DB 859 AGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 908

RESULT 5  
 AAD10117  
 ID AAD10117 standard; cDNA; 1480 BP.

XX AAD10117;

XX 12-Sep-2001 (first entry)

XX Mouse Toso cDNA.

XX Mouse, cytostatic, antiinflammatory; immunoregulatory; tissue integrity;

XX wound healing; immune response; vaccine; cancer; asthma; allergy;

XX cell trafficking; therapy; secreted protein; Fas-induced apoptosis;

XX Toso; ss.

XX Mus sp.

XX Key location/Qualifiers

XX CDS 55..1323

XX /tag= a

XX /product= "Mouse Toso protein"

XX MO200148192-A1.

XX 05-JUL-2001.

XX 21-DEC-2000; 2000MO-N200256

XX 23-DEC-1999, 99US-0171678.

XX 28-NOV-2000; 2000US-0724864

XX (GENE-) GENESTIS RES & DEV CORP LTD.

XX Watson JD, Mullison JG;

XX MPI: 2001-425665/45.

XX P-PSDB: AAY05349.

XX Novel isolated polypeptide useful to isolate corresponding interacting  
 XX proteins or other compounds, to quantitatively determine levels of  
 XX interacting proteins or other compounds, and as therapeutic target  
 XX Claim 1; page 56; 101pp; English.

XX The patent discloses novel polynucleotides and their corresponding  
 XX proteins which play a major role in induction of growth, cell migration  
 XX and proliferation, cell cell interaction and the differentiation of  
 XX tissue-specific cells. These proteins are important in the maintenance















libraries to detect transcripts of a sub transcriptome under a particular biological or pathological state, and so allowing the detection of tissue- and pathology-specific genes such as those genes only expressed in specific tissue under a specific pathological condition; to detect developmental specific genes; and to detect RNA transcripts and splice variants of a transcriptome of a patient suffering from a particular disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from rats, humans and mice, which are used in the exemplification of the present invention.

N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIP0 at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 66 BP, 18 A, 17 C, 13 G, 12 T, 0 other;

Query Match 3.1%; Score 60; DB 24; Length 60;  
Best Local Similarity 100.0%; Fred. No. 5.6e-67;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

249 GTACCTGCTATCCACCACTTCATCAGCCAGATACAGCGCCGACTTCCTGCA 308  
|||||  
1 GTACCTGCTATCCACCACTTCATCAGCCAGATACAGCGCCGACTTCCTGCA 60

RESULT 14  
AA127552/C  
ID AA127552 standard; DNA: 51 BP.  
XX  
AC AA127552;  
XX  
DT 24-JAN-2002 (first entry)  
XX  
DE Human SNP oligonucleotide #760.  
XX  
KM Immunosuppressive; immunostimulatory; antiinflammatory, cytostatic, neuroprotective; antimicrobial; gene therapy; vaccine; amyloid cancer; amyloid protein; angiotensin, angiotensin related protein; cadherin; cyclin; polymerase; oncogene; histone, kinase; colony stimulating factor; complement related protein; cytokine, kinase, cytokine, interleukin, interleukin; G-protein coupled receptor; thioesterase; inflammation; multi-factorial disease; autoimmune disease; infection; nervous system disease; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200147944-A2.  
XX  
PD 05-JUL-2001.  
XX  
PP 28-DEC-2000; 2000WO-US35498.  
XX  
PR 28-DEC-1999; 990S-0173419.  
PP 27-DEC-2000; 2000US-0173419.  
XX  
PA (CUBA-) CUBACHN CORP.  
XX  
PI Shinkels RA, Leach M;  
XX  
DR WPI: 2001-465216/50.  
XX  
PT Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases, oncogenes and histones, useful for diagnosing and treating, e.g. cancer, autoimmune diseases and infections -  
XX  
PS Claim 1; Page 1602, 4143pp; English.  
XX  
CC The present invention relates to oligonucleotides encoding polymorphic variants of proteins related to amylases, amyloid proteins, angiotensin, angiotensin related proteins, cadherin, cyclin, polymerase, oncogenes, histones, kinases, colony stimulating factors, complement related proteins, cytokines, kinases, cytokines, interleukins, interleukins, G-protein coupled receptors and thioesterases. The present sequence is one such oligonucleotide. The oligonucleotides and the peptides encoded

by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate expression of the proteins listed above. Disorders that may be prevented, diagnosed and/or treated include multi-factorial diseases with a genetic component, such as autoimmune diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes, systemic lupus erythematosus and Graves disease), inflammation, cancer (e.g. cancers of the bladder, brain, breast, colon and kidney; leukemia), diseases of the nervous system and an infection of pathogenic organisms.

Sequence 51 BP; 13 A; 7 C; 17 G; 14 T; 0 other;

Query Match 2.7%; Score 51; DB 22; Length 51;  
Best Local Similarity 100.0%; Fred. No. 0.00018;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

525 CTCATATATGCTAGTCTCTTCCAAATTCGTACCCAGATACCAAGTCAGCTG 575  
|||||  
51 CTCATATGCTAGTCTCTTCCAAATTCGTACCCAGATACCAAGTCAGCTG 1

RESULT 15  
AAx84304  
ID AAx84304 standard; DNA: 2320 BP.  
XX  
AC AAx84304;  
XX  
DT 08-SEP-1999 (first entry)  
XX  
DE Human BS135 coding sequence consensus sequence.  
XX  
KM BS135; human; breast tissue; detection; diagnosis; breast cancer; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN W09934017-A1.  
XX  
PD 08-JUL-1999.  
XX  
PP 16-DEC-1998; 96WO-US26918.  
XX  
PR 26-DEC-1997; 97US-0998496.  
XX  
PA (ABB0 ) ABBOTT LAB.  
XX  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;  
PI Gordon J, Grandos EN, Hodges SC, Klass MR, Kratochvil JD;  
PI Roberts-Kapp L, Russell JC, Stroupe SD;  
XX  
DR WPI: 1999-405520/34.  
XX  
PT New isolated BS135 breast tissue polynucleotides useful in the prevention or treatment of breast cancer  
XX  
PS Claim 12; Fig 1; 132pp; English.  
XX  
CC This sequence represents an isolated BS135 breast tissue polynucleotide of the invention. The products can be used for detecting, diagnosing, staging, monitoring, prognosticating, in vivo imaging, preventing or treating, or determining predisposition of diseases or conditions of the breast, such as breast cancer. The products can also be used for drug screening and gene therapy.  
XX  
SV Sequence 4320 BP; 499 A; 760 C; 646 G; 475 T; 0 other;

Query Match 2.4%; Score 46.6; DB 20; Length 2320;  
Best Local Similarity 43.1%; Fred. No. 0.024;  
Matches 223; Conservative 0; Mismatches 294; Indels 0; Gaps 0;

544 CCAATTCATATGCTAGTCTCTTCCAAATTCGTACCCAGATACCAAGTCAGCTG 604  
|||||  
857 CCAATTCATATGCTAGTCTCTTCCAAATTCGTACCCAGATACCAAGTCAGCTG 916



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2002, 02:35:00 : Search time 87 seconds

(Without alignments)  
6736.314 Million cell updates/sec

Title:

US-09-651-150B-1

Perfect score:

1911

Sequence:

1 aaagagtaacagcagctgac . . . ttaactgtctccatcctt 1911

Scoring table:

IDENTITY: 100%  
Gap: 10 0, Expect 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

- 1: Issued\_Patents\_NA.\*
- 2: /cgn2-6/prodata/2/ina/5A\_COMB.seq.\*
- 3: /cgn2-6/prodata/2/ina/5B\_COMB.seq.\*
- 4: /cgn2-6/prodata/2/ina/5C\_COMB.seq.\*
- 5: /cgn2-6/prodata/2/ina/5D\_COMB.seq.\*
- 6: /cgn2-6/prodata/2/ina/5E\_COMB.seq.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1842.4	96.4	2040	US-08-961-564A-1	Sequence 1, Appl
2	676	35.4	1047	US-08-961-564A-3	Sequence 3, Appl
3	563.6	29.5	1480	US-09-724-864-11	Sequence 11, Appl
4	56.2	2.9	7218	US-08-232-463-14	Sequence 14, Appl
5	48.2	2.5	2430	US-09-232-160-16	Sequence 16, Appl
6	43.4	2.3	7218	US-08-232-463-14	Sequence 14, Appl
7	41.8	2.2	1433	US-08-232-463-14	Sequence 14, Appl
8	41.8	2.2	1433	US-08-232-463-14	Sequence 14, Appl
9	40.8	2.1	152331	US-09-199-926-1	Sequence 1, Appl
10	40.6	2.1	7812	US-09-199-926-1	Sequence 1, Appl
11	38.4	2.0	1920	US-09-160-496-4	Sequence 1, Appl
12	37.6	2.0	1984	US-08-937-067-5	Sequence 5, Appl
13	36.6	1.9	936	US-08-416-5090-2	Sequence 4, Appl
14	36.6	1.9	1515	US-08-221-750A-4	Sequence 4, Appl
15	36.6	1.9	7742	US-08-221-750A-1	Sequence 1, Appl
16	36.4	1.9	672	US-09-565-808-3	Sequence 3, Appl
17	36.4	1.9	1044	US-07-975-526-3	Sequence 3, Appl
18	36.4	1.9	1044	US-07-974-409C-425	Sequence 425, App
19	36.4	1.9	1929	PCT-US93-00031-18	Sequence 18, Appl
20	36.4	1.9	1932	PCT-US93-00031-20	Sequence 20, Appl
21	36.4	1.9	1941	PCT-US93-00031-10	Sequence 10, Appl
22	36.4	1.9	1941	PCT-US93-00031-22	Sequence 22, Appl
23	36.4	1.9	2206	PCT-US94-00041-12	Sequence 12, Appl
24	36.4	1.9	2208	PCT-US93-00031-14	Sequence 14, Appl
25	36.4	1.9	2217	PCT-US94-00041-18	Sequence 18, Appl
26	36.4	1.9	2220	PCT-US93-00031-16	Sequence 16, Appl
27	36.4	1.9	2811	US-08-482-073-3	Sequence 3, Appl

#### ALIGNMENTS

28	36.4	1.9	2813	2	US-08-344-155C-99	Sequence 99, Appl
29	36.4	1.9	2813	4	US-09-009-490A-90	Sequence 90, Appl
30	36.4	1.9	3080	4	US-08-482-073-4	Sequence 4, Appl
31	36.4	1.9	3432	4	US-09-423-840-11	Sequence 11, Appl
32	36.2	1.9	6124	4	US-08-213-417B-3	Sequence 3, Appl
33	36.2	1.9	4403765	4	US-09-103-840A-2	Sequence 2, Appl
34	36.2	1.9	4411529	4	US-09-103-840A-1	Sequence 1, Appl
35	36	1.9	5173	1	US-08-232-672-1	Sequence 1, Appl
36	35.8	1.9	289	4	US-08-007-005-17	Sequence 17, Appl
37	35.8	1.9	289	4	US-09-234-796-17	Sequence 17, Appl
38	35.8	1.9	429	1	US-08-463-128-3	Sequence 3, Appl
39	35.8	1.9	429	1	US-08-463-298-3	Sequence 3, Appl
40	35.8	1.9	429	1	US-08-463-339A-3	Sequence 3, Appl
41	35.8	1.9	429	5	PCT-US94-10395-3	Sequence 3, Appl
42	35.8	1.9	18318	1	US-08-414-026A-6	Sequence 6, Appl
43	35.8	1.9	18318	2	US-08-026-022-6	Sequence 6, Appl
44	35.8	1.9	18318	3	US-09-253-682-6	Sequence 6, Appl
45	35.8	1.9	18318	4	US-09-527-657-6	Sequence 6, Appl

#### RESULT 1

US-08-961-564A-1  
Sequence 1, Application US/08961564A

Patent No. 6114515

GENERAL INFORMATION:

APPLICANT: WO, SHUJIAN

APPLICANT: SWEET, KAYMOND

APPLICANT: TRUNER, ALPHEDE

TITLE OF INVENTION: PIGRL-1, A MEMBER OF IMMUNOGLOBULIN

TITLE OF INVENTION: GENE SUPERFAMILY

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: RAINIER & PRESTIA

STREET: P.O. BOX 980

CITY: VALLEY FORGE

STATE: PA

COUNTRY: USA

ZIP: 19482

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,564A

FILING DATE: 30-Oct-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/056,935

FILING DATE: 25-AUG-1997

ATTORNEY/AGENT INFORMATION:

NAME: PRESTIA, PAUL F

REGISTRATION NUMBER: 23,031

REFERENCE/DOCKET NUMBER: GH-70236

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-407-0700

TELEFAX: 610-407-0701

TELEX: 846169

INFORMATION FOR SEQ ID NO. 1.

SEQUENCE CHARACTERISTICS:

LENGTH: 2040 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-961-564A-1

Query Match: 96.4%, Score 1842.4; DB 3; Length 2040;  
Best local Similarity: 99.2%, Prod. No. 0;  
Matches 1894; Conservative 0; Mismatches 11; Gaps 4;









QY 496 ATTACATGCT 506  
1 11111  
DB 1055 CCGCAGCTCC 1045

RESULT 7  
US-09-666-392A-1/C

Sequence 1, Application US/08066392A  
Patent No. 5929040

GENERAL INFORMATION:

APPLICANT: (countries other than U.S.): Royal Children's Hospital

APPLICANT: (U.S. only): George A. WERTHER and

APPLICANT: Christopher J. WRIGHT

TITLE OF INVENTION: A METHOD FOR THE PROPHYLAXIS AND/OR

TITLE OF INVENTION: TREATMENT OF PROLIFERATIVE AND/OR

TITLE OF INVENTION: INFLAMMATORY SKIN DISORDERS

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSER: Brumbaugh, Graves, Donohue & Raymond

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: NY

COUNTRY: U.S.A.

ZIP: 10112-0228

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/666,392A

FILING DATE: 20-AUG-1996

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/00410

FILING DATE: 06-JUL-1995

ATTORNEY/AGENT INFORMATION:

NAME: MacLeod, Janet M.

REGISTRATION NUMBER: 35,263

REFERENCE/DECKET NUMBER: A30625-PCT-USA

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-408-2500

TELEFAX: 212-765-2519

INDEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1433 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 118...1101

OTHER INFORMATION:

NAME/KEY: Coding Sequence

LOCATION: 235...1101

OTHER INFORMATION:

PUBLICATION INFORMATION:

AUTHORS: Hinkert, C., et al.

TITLE: "199109, sequence analysis and expression...

JOURNAL: EMBO J.

VOLUME: 8

ISSUE: 1989

PAGES: 2497-2502

DATE:

US-08-666-392A-1

Query Match 2.2% Score 41.8; DB 2; Length 1433;  
Best Local Similarity 4b.7% Pred. No. 0.019;  
Matches 133; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 841 CCCGACATCCCTGGCCCTTTCTCTGCGACCTTCGCGCGCTGCGTGAAGGAGGCGCT 900

DB 366 CCGGAGAGTTCGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 307

QY 901 TCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960

DB 306 CCGCGGACG 247

QY 961 GAGAGAGTTCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020

DB 246 CAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 187

QY 1021 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080

DB 186 CAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 127

QY 1081 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1125

DB 126 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 82

RESULT 8  
US-09-199-926-1/C

Sequence 1, Application US/09199926  
Patent No. 6284741

GENERAL INFORMATION:

APPLICANT: (countries other than U.S.) Royal Children's Hospital

APPLICANT: (U.S. only): George A. WERTHER and

APPLICANT: Christopher J. WRIGHT

TITLE OF INVENTION: A METHOD FOR THE PROPHYLAXIS AND/OR

TITLE OF INVENTION: TREATMENT OF PROLIFERATIVE AND/OR

TITLE OF INVENTION: INFLAMMATORY SKIN DISORDERS

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSER: Brumbaugh, Graves, Donohue & Raymond

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: NY

COUNTRY: U.S.A.

ZIP: 10112-0228

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/199,926

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/666,392

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: MacLeod, Janet M.

REGISTRATION NUMBER: 35,263

REFERENCE/DECKET NUMBER: A30625-PCT-USA

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-408-2500

TELEFAX: 212-765-2519

INDEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1433 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:



NAME/KEY: Coding Sequence  
 LOCATION: 118...1101  
 OTHER INFORMATION:  
 NAME/KEY: Coding Sequence  
 LOCATION: 118...234  
 OTHER INFORMATION:  
 NAME/KEY: Coding Sequence  
 LOCATION: 235...1101  
 OTHER INFORMATION:  
 PUBLICATION INFORMATION:  
 AUTHORS: Binkert, C., et al.  
 TITLE: Cloning, sequence analysis and expression...  
 JOURNAL: EMBO J.  
 VOLUME: 8  
 ISSUE: 1989  
 PAGES: 2497-2502  
 DATE:  
 US-09-199-926-1

Query Match 2.2%; Score 41.8; DB 4; Length 1433;  
 Best Local Similarity 46.7%; Pred. No. 6.019;  
 Matches 133; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 841 CCGACACACCTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 900  
 DB 366 CCGACACACCTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 307  
 QY 901 TTTAAATATATAAGTCTTTTAAATATATAAGTCTTTTAAATATATAAGT 960  
 DB 306 CCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 247  
 QY 961 CGACAGCTCCCGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1040  
 DB 246 GAACTAAATATAAGTCTTTTAAATATATAAGTCTTTTAAATATATAAGT 187  
 QY 1021 CCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 1080  
 DB 186 CCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 127  
 QY 1081 CCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 1125  
 DB 126 CCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 82

RESULT 9  
 US-09-128-155-16  
 Sequence 16, Application US/09128155  
 Patent No. 6117654  
 GENERAL INFORMATION:  
 APPLICANT: Pan, Yang  
 TITLE OF INVENTION: NOVEL MOLECULES OF TAMBO-77 RELATED PROTEIN FAMILY  
 FILE REFERENCE: 09404/052001  
 CURRENT APPLICATION NUMBER: US/09/128,155  
 EARLIER FILING DATE: 1998-07-02  
 EARLIER APPLICATION NUMBER: US 60/054,646  
 EARLIER FILING DATE: 1997-08-04  
 NUMBER OF SEQ ID NOS: 18  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO: 16  
 LENGTH: 152331  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (1)...(152331)  
 OTHER INFORMATION: n - A,T,C or G  
 US-09-128-155-16

Query Match 2.1%; Score 40.8; DB 3; Length 152331;  
 Host Local Similarity 47.8%; Pred. No. 0.62;

Matches 117; Conservative 0; Mismatches 128; Indels 0; Gaps 0;  
 QY 915 CCGCTTCTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 974  
 DB 21916 CCGCTTCTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21975  
 QY 975 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 1034  
 DB 21976 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 22035  
 QY 1035 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 1094  
 DB 22036 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 22095  
 QY 1095 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 1154  
 DB 22096 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 22155  
 QY 1155 AGACG 1159  
 DB 22156 GCGGCT 22160

RESULT 10  
 US-09-368-590-1/c  
 Sequence 1, Application US/09368590  
 Patent No. 6187563  
 GENERAL INFORMATION:  
 APPLICANT: Solimena, Michele  
 TITLE OF INVENTION: INTERACTING POLYPEPTIDES FOR  
 TITLE OF INVENTION: AUTOANTIGENS OF AUTOIMMUNE DISEASES  
 FILE REFERENCE: 101918-200 (OCR-941)  
 CURRENT APPLICATION NUMBER: US/09/368,590  
 CURRENT FILING DATE: 1999-08-04  
 EARLIER APPLICATION NUMBER: 60/095,657  
 EARLIER FILING DATE: 1998-08-07  
 NUMBER OF SEQ ID NOS: 8  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO: 1  
 LENGTH: 7812  
 TYPE: DNA  
 ORGANISM: Human  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)...(6879)  
 NAME/KEY: unsure  
 LOCATION: (100)...(104)  
 NAME/KEY: unsure  
 LOCATION: (1021)...(1023)  
 NAME/KEY: unsure  
 LOCATION: (2266)....(2268)  
 US-09-368-590-1

Query Match 2.1%; Score 40.6; DB 4; Length 7812;  
 Best Local Similarity 50.5%; Pred. No. 0.12;  
 Matches 97; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 924 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 983  
 DB 1164 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 1105  
 QY 984 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 1043  
 DB 1104 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 1045  
 QY 1044 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 1104  
 DB 1044 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 985  
 QY 1104 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 1115  
 DB 984 GCGGCTAACTGGGCTTTCCTGCGACCTCTGGGGCTGCTGCAAAAGGCGCT 973





Fri Mar 14 13:12:22 2003

us-09-651-150b-1.rni

Page 10

Search completed: November 24, 2002, 04:46:29  
Job time : 216 secs

---



[illegible]

RESULT 2  
US-09-215-652-16

```

Sequence: 16/ Application US/09/215652
Patent No. US20020045165A1
GENERAL INFORMATION:
APPLICANT: Patricia Billing Medel
APPLICANT: Maurice Cohen
APPLICANT: Tracey L. Colpitts
APPLICANT: Paula N. Friedman
APPLICANT: Julian Gordon
APPLICANT: Edward N. Granados
APPLICANT: Steven C. Hodges
APPLICANT: Michael R. Kloss
APPLICANT: Jon D. Kratochvil
APPLICANT: Lisa Roberts-Rapp
APPLICANT: John C. Russell
APPLICANT: Stephen D. Stroupe
TITLE OF INVENTION: Reagents and Methods Useful for Detecting Diseases of the
TITLE OF INVENTION: Breast
FILE REFERENCE: 6192 US P1
CURRENT APPLICATION NUMBER: US/09/215,652
CURRENT FILING DATE: 1998-12-16
EARLIER APPLICATION NUMBER: US 08/998,496
EARLIER FILING DATE: 1997-12-26
NUMBER OF SEQ. ID NOS. 48
SOFTWARE: FASTSeq for Windows Version 3.0
SEQ. ID NO 16
LENGTH: 2907
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: base_polymorphism
LOCATION: 11
OTHER INFORMATION: /note = " represents an a or g or t or c polymorphism
OTHER INFORMATION: this position
US-09-215-652-16

```

Query Match:	248	Score 46.6	DB 10	Length 2907
Host local Similarity	43.1%		pred No 0.0012	
Matches 223	Conservative 0	Mismatches 294	Indels 0	Caps 0

[illegible]

### RESULT 3

US-09-880-107-2344/c  
Sequence 2344, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Schert, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIORITY FILING DATE: 2000-06-14  
PRIORITY FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIORITY FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2344  
LENGTH: 1387  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US:299,914.2,981A1 M354110  
US-09-880-107-2344

Query Match	2.28;	Score 41.8;	DB 10;	Length 1387;
Best local Similarity	46.78;	pred. NO. 0.922;		
Matches 133;	Conservative	0;	Mismatches 152;	Indels 0;
				Gaps 0;

[illegible]



Query Match	2.28;	Score 41.4,	DB 12;	Length 12425,
Best local Similarity	50.88;	Pred. No. 0.14;		
Matches	99,	Conservative	0;	Mismatches 96;
				Indels 0;
				Gaps 0;

[illegible]

RESULT 9  
US-09-764-853-104/C  
; Sequence 104, Application US/09764853





```

PRIOR FILING DATE: 1999-06-09
PRIOR APPLICATION NUMBER: 60/144791
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/169495
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/175481
PRIOR FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: 60/191007
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/199397
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/380139
PRIOR FILING DATE: 1998-08-25
PRIOR APPLICATION NUMBER: 09/311832
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/380137
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380138
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380142

Query Match 2.1% Score 39.8, DB 12, Length 1505;
Best Local Similarity 52.0%; Freq. No. 0.095;
Matches 89, Conservative 0, Mismatches 82, Indels 0, Gaps 0;

UY 985 GTGCGCCGACCGCCGCTCCCAAAACAGATCTACAGCCGCTGCGCCGCGCTGCTG 1044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 252 GTTGCCGACATGATGGGGGCGAGCGCCAGCCATCTGATGCTTGGTGGAGCGAGGCGCGG 193

UY 1045 AGCGGAGCGCTGACAGGACACAGGGGAGGCGCCGCTGCGCCGCGCGCGCGCGCGCG 1104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 192 CACGGCGCGCCCGCATGCGCCCGGAGCGCCCGCTGACAGCGCGCGCGCGCGCGCGCG 133

UY 1105 GCGCGCGCTGCAAGTGTGTGAACTCTCTGCTGATCTGCTGCTGCTGCTGCTGCTG 1155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132 CTCGATGCGCGCGCTGCGCATGGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82

RESULT 11
US-10-052-586-377/c
: Sequence 377, Application US/10052586
: Patient No. US20020127584A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/052, 586
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28

```

[illegible]



US-09-795-693-7

Query Match 2 1% Score 39.6; PR 10; Length 8195;  
Best Local Similarity 52.4%; Pred. No. 0.36;  
Matches 87; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 944 CTGAGATGCCCCCGGAGAGCTCCAGAGCCCCCGGCGCCGCGCCGCGC 1003  
DB 226 GGGATGAGATTTTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 157  
QY 1004 CAAAGACACATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1063  
DB 166 CAGAGCAG 107  
QY 1064 GAG 1109  
DB 106 GAG 61

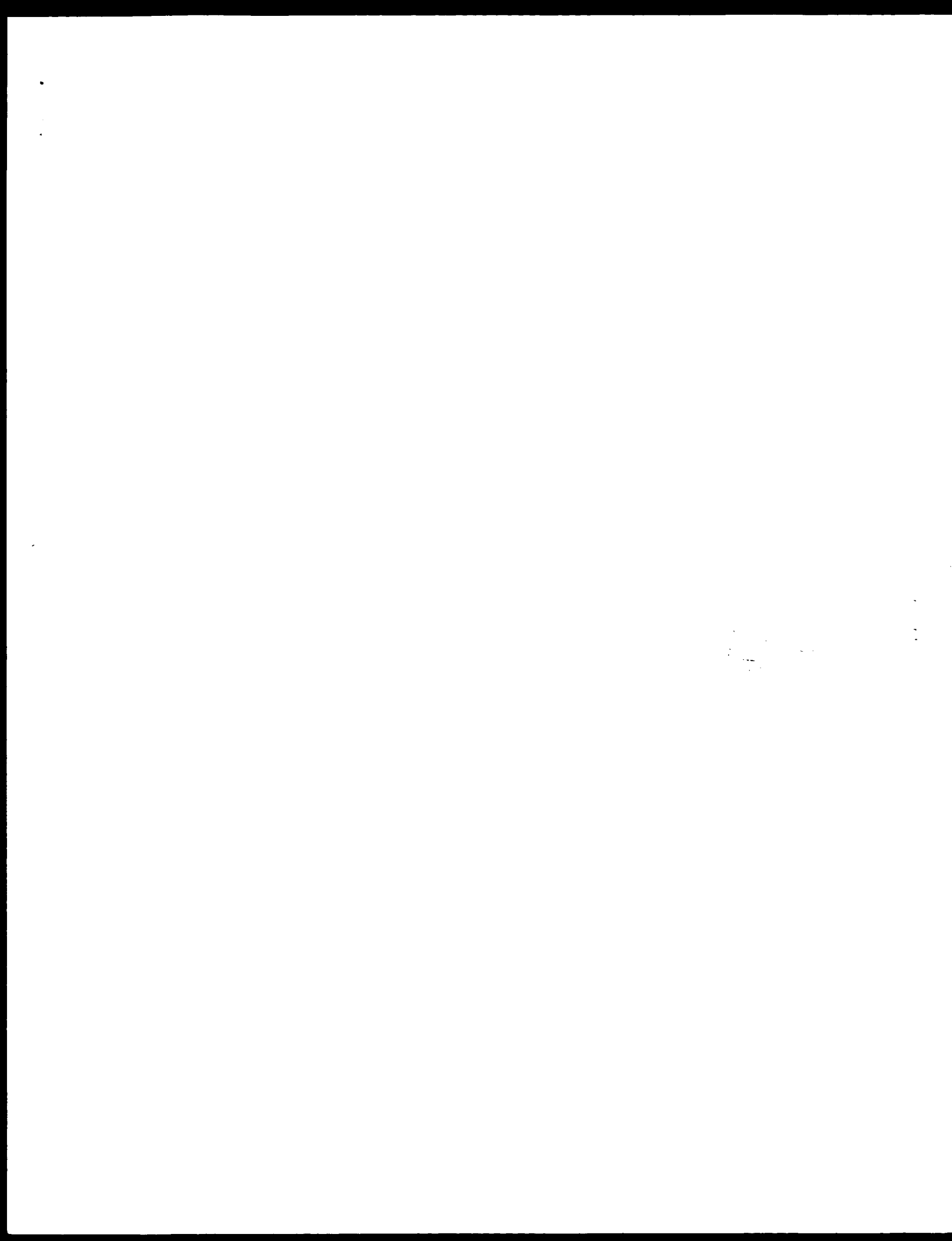
RESULT 15

US-09-729-674-153/C  
Sequence 153, Application US/09729674  
Patent No. US20010039335A1  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallee, Edward R.  
APPLICANT: Collins-Raele, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Agostino, Michael J.  
APPLICANT: Steindler II, Robert J.  
APPLICANT: Spaulding, Vikki  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Pechel, Kim  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: 6055-64X  
CURRENT APPLICATION NUMBER: US/09/729,674  
CURRENT FILING DATE: 2000-12-04  
PRIORITY FILING DATE: 09/539,330  
PRIORITY FILING DATE: 2000-03-30  
NUMBER OF SEQ ID NOS: 283  
SOFTWARE: Patentia Ver. 2.0  
SEQ ID NO 153  
LENGTH: 3343  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-729-674-153

Query Match 2.0%; Score 38.8; DB 10; Length 3343;  
Best Local Similarity 48.2%; Pred. No. 0.34;  
Matches 109; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 914 GCTTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 973  
DB 327 GCGATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 268  
QY 974 AGGCG 1033  
DB 267 GCGATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 208  
QY 1014 GCGATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1093  
DB 207 CTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148  
QY 1094 CTTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1159  
DB 147 CAGTGGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 102

Search completed: November 24, 2002, 05:54:03  
Job time : 238 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd

GM nucleotide multiple search, using SW mode

Run on November 24, 2002, 02:34:40, Search time 2445 Seconds

(without alignments)  
12658.313 Million cell updates/sec

Title: US-09-651-150b-1  
Perfect score: 1911  
Sequence: 1 daagagatgaagcagcagcgttc  
tctactctgtctctatctctt 1911

Scoring table:  
Gapop 10 0 Gapext 1 0

Searched: 1615400b seqs, 800744437b residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: EST:  
1: em\_estb1:  
2: em\_estb1:  
3: em\_estb1:  
4: em\_estb1:  
5: em\_estb1:  
6: em\_estb1:  
7: em\_estb1:  
8: em\_estb1:  
9: em\_estb1:  
10: em\_estb1:  
11: em\_estb1:  
12: em\_estb1:  
13: em\_estb1:  
14: em\_estb1:  
15: em\_estb1:  
16: em\_estb1:  
17: em\_estb1:  
18: em\_estb1:  
19: em\_estb1:  
20: em\_estb1:  
21: em\_estb1:  
22: em\_estb1:  
23: em\_estb1:  
24: em\_estb1:  
25: em\_estb1:  
26: em\_estb1:  
27: em\_estb1:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	879	46.0	896	9	AL568724
2	878	45.9	904	14	AL568724
3	821.4	43.0	894	14	AL568724
4	811.4	42.5	895	12	AL568724
5	773.8	40.5	928	12	AL568724
6	750.8	39.3	936	14	AL568724

Result No.	Score	Query Match	Length	DB ID	Description
7	727.8	38.1	763	13	AL568724
8	714.4	37.4	1070	14	AL568724
9	704.6	36.9	912	14	AL568724
10	694.6	36.4	842	12	AL568724
11	682.2	35.7	838	12	AL568724
12	657.4	34.4	765	12	AL568724
13	647.8	33.9	711	9	AL568724
14	645.2	33.8	915	12	AL568724
15	641.4	33.6	696	13	AL568724
16	636.2	33.3	670	13	AL568724
17	627.4	32.8	730	13	AL568724
18	617	32.3	815	12	AL568724
19	610	31.9	994	12	AL568724
20	583	30.5	667	12	AL568724
21	572.2	29.9	824	12	AL568724
22	565	29.6	673	12	AL568724
23	564	29.5	614	12	AL568724
24	563.6	29.5	1492	11	AL568724
25	556	29.1	974	12	AL568724
26	522	27.3	578	10	AL568724
27	515.4	27.0	549	10	AL568724
28	485.6	25.4	549	10	AL568724
29	478.6	25.0	907	12	AL568724
30	464.6	24.5	844	12	AL568724
31	462.8	24.2	508	10	AL568724
32	458.8	24.0	524	10	AL568724
33	458.2	24.0	507	12	AL568724
34	456.2	23.9	507	12	AL568724
35	449.4	23.5	503	10	AL568724
36	446	23.3	525	10	AL568724
37	432	22.6	510	10	AL568724
38	431	22.6	509	9	AL568724
39	431	22.6	511	9	AL568724
40	431	22.6	511	10	AL568724
41	431	22.6	513	10	AL568724
42	428.2	22.4	475	13	AL568724
43	421.6	22.1	459	10	AL568724
44	418.2	21.9	456	9	AL568724
45	418	21.9	488	9	AL568724

## ALIGNMENTS

RESULT 1  
AL568724/c  
LOCUS: AL568724  
DEFINITION: AL568724  
ACCESSION: AL568724  
VERSION: AL568724  
KEYWORDS: EST.  
SOURCE: human.  
ORGANISM: Homo sapiens  
REFERENCE: Li, W. B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE: Full-length cDNA libraries and normalization  
JOURNAL: Unpublished (2001)  
COMMENT: Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

## FEATURES

1. 896  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="US09-651-150b-1"  
/clone\_lib="LTI-FL002\_P11"  
/lab\_host="DH10B"  
/note="Organ: Placenta, Vector: pCMVSPORT1 6; 1st strand  
cDNA was primed with a NotI oligo(dT) primer. Five prime

end enriched, double-stranded cDNA was ligated with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was constructed by life technologies. Contact: Feng liang life technologies, a division of invitrogen 9800 medical center drive rockville Maryland 20850, USA Fax: (1) 301 410 8471 Email: lian@life technologies.com tpl: lian@life technologies.com

	http://fulllength.invitrogen.com"			
BASH COUNT	164	a	217	c
ORIGIN	329	g	182	t
				4 others

Query Match	46.08;	Score 879;	DB 9;	length 896;
Best Local Similarity	99.28;	Pred. No. 5.7e-219;		
Matches 890; Conservative	3;	Mismatches 3;	Indels 1;	Gaps 1

[illegible]

Db	56	OTMA-TTAT-ATTATGTT-9007-ME-AH86FTGGXGLAAADICT	1
RESULT 2			
LOCUS	AL541079	904 bp	mRNA linear EST 16-FEB-2001
DEFINITION	AL541079 LTI_FL002_F11 Homo sapiens cDNA clone GS069057c10 5 prime		
ACCESSION	AL541079		mRNA sequence.
VERSION	AL541079.1	G1:12871804	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Akayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Euteria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	Ll W.H., Gruber C., Jossoc J. and Polayes D.		
JOURNAL	Full-length cDNA libraries and normalization		
COMMENT	Unpublished (2001)		
	Contact: Genoscope		

Email: [segrete@genoscope.cns.fr](mailto:segrete@genoscope.cns.fr), Web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/cellone="GSDPE0057101"
/cellone.lib="l11_FL002_pl1"
/lab_host="DH10B"
/node="Organ: Placenta; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9400 Medical Center Drive Pockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
f.liang@lifetech.com URL :
http://fulllength.invitrogen.com"
232 a 260 c 230 g 180 t 2 others

```

Query Match	45.38;	Score 878;	DB 9,	Length 904;
Best Local Similarity	99.3%;	Pred. No. 1e-218;		
Matches 900;	Conservative	2;	Mismatches	2;
			Indels	2;
			Gaps	2

[illegible]







FEATURES  
source location/Qualifiers  
1..928  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="NH\_MGC\_48"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (Phage-resistant)"  
/note="Organ: B-cells; Vector: pORF7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GCCACGAC(C). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH-MGC library."

BASE COUNT 241 a 261 c 253 g 171 t 2 others

Query Match 40.5%; Score 773.8; DB 13; Length 928;  
Best Local Similarity 94.5%; Pred. No. 2.1e-191;  
Matches 876; Conservative 0; Mismatches 44; Indels 7; Gaps 7;

43 CTTCGATGACCTTCGACCTAGACGACATGACCTTCGCTTCGACCTTACTT 102  
|||||  
2 CTTCGATGATTTTATCTGATGATGATGATGATGATGATGATGATGATGAT 60  
|||||  
103 CCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 162  
|||||  
61 CCGCCGACGATCG 120  
|||||  
163 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 222  
|||||  
121 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
|||||  
223 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 282  
|||||  
181 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
|||||  
283 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 342  
|||||  
241 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
|||||  
343 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 402  
|||||  
301 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
|||||  
403 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 462  
|||||  
361 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
|||||  
463 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 522  
|||||  
421 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
|||||  
523 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 582  
|||||  
481 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
|||||  
583 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 642  
|||||  
541 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 599  
|||||  
643 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 702  
|||||  
600 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 658  
|||||  
703 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 762  
|||||  
659 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 718  
|||||  
763 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 821

|||||  
Db 719 - ANCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777  
|||||  
Qy 822 AAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 880  
|||||  
Db 778 CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837  
|||||  
Qy 881 CTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 939  
|||||  
Db 838 CTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897  
|||||  
Qy 940 GCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 966  
|||||  
Db 898 GCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 924

RESULT 6  
BO707120 936 bp mRNA linear EST 16 JUL 2002  
LOCUS AGNCOURT\_8475243 NH\_MGC\_113 Homo sapiens cDNA clone IMAGE:6301366  
DEFINITION 5', mRNA sequence.  
ACCESSION BO707120  
VERSION BO707120.1 GI:21846019  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 936)  
NIH-MGC <http://mgi.mcg.mcg.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
CONTACT Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Dr. Mark Watson  
cDNA Library Preparation: Rubin Laboratory  
cDNA library Arrayed by: The I.M.A.G.E. Consortium (ILUNI)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plates: IL02517 row: b column: 23  
High quality sequence stop: 624.  
Location/Qualifiers  
1..936  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="NH\_MGC\_113"  
/lab\_host="DH10B (Phage-resistant)"  
/note="Organ: spleen; Vector: pTZ19; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GCCACGAC(C). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH-MGC library."

BASE COUNT 246 a 279 c 234 g 197 t

Query Match 39.3%; Score 750.8; DB 14; Length 936;  
Best Local Similarity 97.5%; Pred. No. 2.3e-185;  
Matches 794; Conservative 0; Mismatches 17; Indels 3; Gaps 3;

5 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 64  
|||||  
Db 57 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 116  
|||||  
Qy 65 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 124  
|||||  
Db 117 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 176  
|||||















QY 673 GCTCCGAACTTCCTGATTCATACAGCTGAAA 708  
 DB 660 GCTCCGAACTTCCTGATTCATACAGCTGAAA 693

RESULT 15  
 H1836254

LOCUS B1836254 698 bp mRNA linear EST 04-OCT-2001  
 DEFINITION 603083927F1 NIH\_MGC\_120 Homo sapiens cDNA clone IMAGE 5223150 5',  
 mRNA sequence.

ACCESSION B1836254  
 VERSION B1836254  
 KEYWORDS EST,  
 SOURCE human,  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 698)  
 AUTHORS NIH-MGC http://mgi.mcg.mil.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-st@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
 DNA Sequencing by: Tcycle Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
 http://image.llnl.gov  
 plate: L1AM11561 row: c column: 07  
 High quality sequence stop: 696.  
 Location/Qualifiers

FEATURES  
 source  
 1..698  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5223150"  
 /clone\_lib="NIH\_MGC\_120"  
 /lab\_host="DH10B"  
 /note="Organ: pooled pancreas and spleen; Vector:  
 pCMV-Sport6; Site\_1: NotI; Site\_2: KcoRV (destroyed); RNA  
 source anonymous pool of spleen and pancreas from 28 yo  
 male. Library is oligo-dT primed and directionally cloned  
 (EcoRV site is destroyed upon cloning). Average insert  
 size 1.5 kb, insert size range 1.2-5 kb. Library is  
 normalized and enriched for full-length clones and was  
 constructed by C. Gruber (Invitrogen). Research Genetics  
 tracking code 025. Note: this is a NIH MGC Library."  
 BASE COUNT 185 a 195 c 174 g 143 t 1 others  
 ORIGIN

Query Match 33.6% Score 641.4; DB 13; Length 698;  
 Best Local Similarity 98.4%; Pred. No. 8.6e-157;  
 Matches 689; Conservative 0; Mismatches 7; Indels 4; Caps 4;

QY 32 CTTTAAAGGATTTTATGAGAAATTGATATTAAGAGAAATGATTTTGGTTTG 91  
 DB 1 CTTTAAAGGATTTTATGAGAAATTGATATTAAGAGAAATGATTTTGGTTTG 59  
 QY 92 CCACTTACCTCCGCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 151  
 DB 60 CCACTTACCTCCGCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119  
 QY 152 GAGGAGATGATATGATATGATATGATATGATATGATATGATATGATATG 211  
 DB 120 GAGGAGATGATATGATATGATATGATATGATATGATATGATATGATATG 179  
 QY 212 TATCTATGATATGATATGATATGATATGATATGATATGATATGATATGAT 271  
 DB 180 TATCTATGATATGATATGATATGATATGATATGATATGATATGATATGAT 239  
 QY 272 TTATCAAGGAGAGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 331

DB 240 TTATCAAGGAGAGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 299  
 QY 332 TTCTATAG-TGGAGGTAAACAGATTAATAGAAATGATATGATATGATATG 390  
 DB 300 TTCTATAGATGAGGTAAACAGATTAATAGAAATGATATGATATGATATG 359  
 QY 391 GAGGAGATGATATGATATGATATGATATGATATGATATGATATGATATG 450  
 DB 360 GAGGAGATGATATGATATGATATGATATGATATGATATGATATGATATG 419  
 QY 451 GAGGAGATGATATGATATGATATGATATGATATGATATGATATGATATG 510  
 DB 420 GAGGAGATGATATGATATGATATGATATGATATGATATGATATGATATG 479  
 QY 511 TTCTATAGATGAGGTAAACAGATTAATAGAAATGATATGATATGATATG 570  
 DB 480 TTCTATAGATGAGGTAAACAGATTAATAGAAATGATATGATATGATATG 539  
 QY 571 AGCTCAAG 630  
 DB 540 AGCTCAAG 598  
 QY 631 GAGGAGATGATATGATATGATATGATATGATATGATATGATATGATATG 689  
 DB 599 GAGGAGATGATATGATATGATATGATATGATATGATATGATATGATATG 658  
 QY 690 GAGGAGATGATATGATATGATATGATATGATATGATATGATATGATATG 759  
 DB 659 GAGGAGATGATATGATATGATATGATATGATATGATATGATATGATATG 698

Search completed: November 24, 2002, 04:42:48  
 Job time: 2453 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd

OM protein - protein search, using SW model

Run on: November 24, 2002, 06:10:41, Search time 22 Seconds

(Without alignments)  
1704.201 Million cell updates/sec

Title: US-09-651-150b-2  
Perfect score: 2055  
Sequence: 1 MDPWIMPIYPLIPVSCAIFIL HQPAMMPDSINIVINVA 390

Scoring table: R/SIM62  
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134122 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: PIR-73:  
1: PIR1:  
2: PIR2:  
3: PIR3:  
4: PIR4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No	Score	Query Match	Length	FM	IR	Description
1	185	9.0	769	1	0.00185	secretory component
2	182	8.9	757	1	0.00182	secretory component
3	182	8.9	757	1	0.00182	secretory component
4	182	8.9	757	1	0.00182	secretory component
5	166.5	8.1	773	1	0.00166	secretory component
6	117	5.7	1106	1	0.00117	transforming protein
7	113.5	5.5	1415	1	0.00113	immediate-early protein
8	112.5	5.5	224	1	0.00112	interferon- $\gamma$
9	112	5.5	2084	1	0.00112	proteasome multi-domain
10	111.5	5.4	390	1	0.00111	glycoprotein 1 protein
11	111.5	5.4	392	1	0.00111	protein C46E10.3
12	111	5.4	801	1	0.00111	hypothetical protein
13	110.5	5.4	2339	1	0.00110	omega-conotoxin-se
14	107.5	5.2	528	1	0.00107	conserved hypothetical
15	107	5.2	1791	1	0.00107	hypothetical protein
16	106.5	5.2	460	1	0.00106	hypothetical protein
17	106	5.2	496	1	0.00106	hypothetical protein
18	103	5.0	478	1	0.00103	hypothetical protein
19	103	5.0	526	1	0.00103	hypothetical protein
20	103	5.0	700	1	0.00103	hypothetical protein
21	102.5	5.0	1008	1	0.00102	SEC14 protein homolog
22	102	5.0	530	1	0.00102	follicular receptor
23	101.5	4.9	166	1	0.00101	heavy chain protein
24	101.5	4.9	456	1	0.00101	proanthranilate N-
25	101.5	4.9	627	1	0.00101	19 mu chain precursor
26	101.5	4.9	2152	1	0.00101	hypothetical protein
27	101.5	4.9	2237	1	0.00101	N-type calcium channel
28	101.5	4.9	3530	1	0.00101	unconventional myosin
29	101	4.9	475	1	0.00101	gamma-2b chain

30	100.5	4.9	504	2	0.00100	transcription factor
31	100.5	4.9	876	1	0.00100	regulatory protein
32	100.5	4.9	1616	2	0.00100	gene APRI protein
33	100	4.9	384	1	0.00100	dual specificity protein
34	99.5	4.8	2274	1	0.00099	adenomatous polyp
35	99	4.8	430	2	0.00099	B lymphocyte antigen
36	98.5	4.8	116	2	0.00098	transit protein
37	98.5	4.8	847	1	0.00098	H cell adhesion protein
38	98	4.8	554	1	0.00098	macrophage colony
39	98	4.8	1021	2	0.00098	connectin/tillo-
40	97.5	4.7	368	2	0.00097	hypothetical protein
41	97.5	4.7	870	2	0.00097	extensin homolog
42	97	4.7	1013	2	0.00097	hypothetical protein
43	97	4.7	1649	1	0.00097	delta-(1-alpha-anti
44	96.5	4.7	1060	2	0.00096	acrosomal protein
45	96.5	4.7	2245	2	0.00096	hypothetical protein

## ALIGNMENTS

### RESULT 1

Secretory component precursor - rat

N:Alternate names: poly Ig receptor, polymeric immunoglobulin receptor

N:Contains: free secretory component, transmembrane secretory component

C:Species: Rattus norvegicus (Norway rat)

C>Date: 07-Sep-1990 #sequence\_revision 2#Aug-1996 #ext\_change 16-10-1999

C:Accession: S05407; S54731

R:Hanley, G.; Blake, B.; Bhaghetta, P.; Lutz, J.P.; Stanley, K.K.

PNRS Lett. 254, 177-183, 1989

A>Title: Intracellular targeting signals of polymeric immunoglobulin receptors are h

A:Reference number: S05407; MUID:89378226; PMID:2776882

A:Accession: S05407

A:Molecule type: mRNA

A:Residues: 1769 (BAN)

A:Cross-references: EMBL:X15741; NID:956464; PIR:NCAA3728.1; PIR:956465

C:Complex: monomeric as a transmembrane receptor or free in mucosal secretions; heter

amers; hetero-22-mer composed of one chain of secretory component, one chain of immun

C:Keywords: duplication; glycoprotein; immunoglobulin receptor; phosphoprotein; trans

F:1-18/Domain: signal sequence #status predicted <SIM>

F:19-769/Product: transmembrane secretory component #status predicted <MAIM>

F:119-582/Product: free secretory component #status predicted <MAIF>

F:33-112/Domain: immunoglobulin homology <IM1>

F:145-222/Domain: immunoglobulin homology <IM2>

F:253-336/Domain: immunoglobulin homology <IM3>

F:363-442/Domain: immunoglobulin homology <IM4>

F:477-548/Domain: immunoglobulin homology <IM5>

F:644-666/Domain: transmembrane #status predicted <IM6>

F:667-769/Domain: intracellular #status predicted <IM7>

F:40-110.56/56-152.420.370-440.394.484.541.498-505/Dissulfide bonds: #status predi

F:90.135-471/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:488-522/Dissulfide bonds: (in Ig-unbound form) #status predicted

F:522/Binding site: cysteine (Cys) (covalent) (in Ig bound form) #status predicted

F:678/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 9.8% Score 185, DE 1; Length 769;

Best Local Similarity 34.58; Pred. No. 110-05;

Matches 41; Conservative 21; Mismatches 41; Indels 16; Gaps 4;

QY 30 GGSVTKPLPPE---MIVPIYL/PEMAGSGTGTIVSTNFIKAEYKGPVTLKQYPRKN 85  
DB 33 GNSVITCYPPDTSVNPHTPKYKWPQGA-NVVCATLISNRYLSKEYSGFASLNPENNS 91  
QY 86 LNVIVTQILTESDGVVACGACAMNDKCKTQKVLINVSFHPDSMHQPMPTKMKHLL 144  
DB 92 TVVINAIHLTDETSYKQGLA-TTNPLPFDVSLV-----SQVPEFNTDTHV 139  
RESULT 2  
S48841

secretory component precursor - bovine

N.A. Interact. names: poly Ig receptor; polymeric immunoglobulin receptor  
N. Contain: free secretory component; transmembrane secretory component

C. Species: Bos primigenius laurus (cattle)

C. Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #excl\_change 10-Sep-1999

C. Accession: S48841

R. Vermeer, H. J. W. M. de Boer, H. A. J. Verheij, M. P. H.

submitted to the EMBL Data Library, September 1994

A. Description: The cloning, tissue specific expression and interspecies sequence compari

A. Reference number: S48841

A. Accession: S48841

A. Status: preliminary

A. Molecule type: mRNA

A. Residues: 1-757 <RES>

A. Cross-references: EMBL:X81371; NID:9563340; PDB:1A557136.1; PDB:9563341

C. Keywords: secretory component; immunoglobulin homology

C. Keywords: duplication; glycoprotein; immunoglobulin receptor; phosphoprotein; transcript

P1-18/Domain: signal sequence #status predicted <SIG>

P19-757/Product: transmembrane secretory component #status predicted <MATM>

P19-757/Product: free secretory component #status predicted <MAT>

P13-112/Domain: immunoglobulin homology <IM1>

P145-222/Domain: immunoglobulin homology <IM2>

P250-328/Domain: immunoglobulin homology <IM3>

P365-444/Domain: immunoglobulin homology <IM4>

P476-547/Domain: immunoglobulin homology <IM5>

P631-653/Domain: transmembrane #status predicted <TM>

P654-757/Domain: intracellular #status predicted <INT>

F140-110, 56-64, 156-173, 257-324, 471-479, 490-440, 384-394, 481-543, 495-502/Disulfide

F183-420, 468/Binding site: carbohydrate (Asn) (covalent) #status predicted

F1665/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match

Best Local Similarity 8.9% Score 182; DB 1; Length 757;

Best Local Similarity 34.1% Pred. No. 1 Re-05;

Matches 42; Conservative 21; Mismatches 44; Indels 16; Gaps 4;

17 LRLLEVEKVEKEL-----CCSVTKKCHP-----EMHVRVYICREMAAGCTCCGCV 62

10 LAIFPVSVSKSP1PCHREVSVKGSVSIKCYPTPTSVNKRTRKWKCOCA-QGRCTTLI 68

63 STTNFKAEYKGPVTLKAYPRKFLVEVLTGLESQSVAVAGAMNTDPRKTKVTLNV 122

69 SSRGVSDYDVGKRNILNFPSCIFVVDIISHLTHKDSCKYKCGICLIS-RLNFDVSLV 127

QY 123 HSE 125

DB 128 SOD 130

RESULT 3

145956

polymeric immunoglobulin receptor - bovine

C. Species: Bos primigenius laurus (cattle)

C. Date: 19-Dec-1997 #sequence\_revision 19-Dec-1997 #excl\_change 23-Jul-1999

C. Accession: 145956

R. Kraljci, M. A. Kraljci, P. J. M. Verheij, O. J. Ruyter, S.

DNA Cell Biol 14, 251-256, 1996

A. Title: Cloning and characterization of two forms of bovine polymeric immunoglobulin re

A. Reference number: 145956; MIMD-9418064; PMID-7860445

A. Accession: 145956

A. Status: preliminary; translated from GH/EMH/70004

A. Molecule type: mRNA

A. Residues: 1-757 <RES>

A. Cross-references: EMBL:U4797; NID:9388279; PDB:1A557136.1; PDB:9488280

C. Superfamily: secretory component; immunoglobulin homology

F145-222/Domain: immunoglobulin homology <IM1>

Query Match 8.9% Score 182; DB 2; Length 757;

Best Local Similarity 34.1% Pred. No. 1 Re-05;

Matches 42; Conservative 21; Mismatches 44; Indels 16; Gaps 4;

17 LRLLEVEKVEKEL-----CCSVTKKCHP-----EMHVRVYICREMAAGCTCCGCV 62

10 LAIFPVSVSKSP1PCHREVSVKGSVSIKCYPTPTSVNKRTRKWKCOCA-QGRCTTLI 68

63 STTNFKAEYKGPVTLKAYPRKFLVEVLTGLESQSVAVAGAMNTDPRKTKVTLNV 122

69 SSRGVSDYDVGKRNILNFPSCIFVVDIISHLTHKDSCKYKCGICLIS-RLNFDVSLV 127

QY 123 HSE 125

DB 128 SOD 130

RESULT 4

ORHUGS

secretory component precursor [validated] - human

N. Alternate names: poly Ig receptor; polymeric immunoglobulin receptor

N. Contain: free secretory component; transmembrane secretory component

C. Species: Homo sapiens (man)

C. Date: 28-Aug-1985 #sequence\_revision 23-Aug-1996 #excl\_change 08-Dec-2000

C. Accession: A46537; A55284; 138115; A32263; S38978; S13453; A02112

R. Kraljci, P. J. Kraljci, D. J. Tasken, K. J. Brandtzaeg, P.

Int. J. Immunol. 22, 2309-2315, 1992

A. Title: Molecular cloning and exon-intron mapping of the gene encoding human transme

A. Reference number: A46537, MIMD:92387236, PMID:1355431

A. Accession: A46537

A. Status: not compared with conceptual translation

A. Molecule type: DNA

A. Residues: 1-764 <RES>

A. Cross-references: EMBL:54345; NID:945007; EMBL:AA02203.1; PDB:9238236

A. Experimental source: leukocytes

A. Note: sequence extracted from NDBI backbone (NCBI:62403)

R. Kraljci, P. J. Kraljci, K. H. J. Geurts van Kessel, A. H. J. Olaisen, B. J. Brandtzaeg, P.

Hum. Genet. 87, 642-648, 1991

A. Title: The human transmembrane secretory component (poly Ig receptor) - molecular cl

A. Reference number: A55284; MIMD:92039621; PMID:1682231

A. Accession: A55284

A. Molecule type: mRNA

A. Residues: 1-764 <RES>

A. Cross-references: EMBL:54345; NID:945007; EMBL:AA02203.1; PDB:9238236

A. Experimental source: colonic adenocarcinoma cell line

A. Note: sequence extracted from NDBI backbone (NCBI:62403)

F. Piskurich, J. F. France, J. A. Tamer, C. M. Wallmer, C. A. Kaelzel, C. S. Kaelzel,

Mol. Immunol. 30, 413-421, 1993

A. Title: Interferon-gamma induces polymeric immunoglobulin receptor mRNA in human int

A. Reference number: 138115, MIMD:94205018, PMID:8455639

A. Accession: 138115

A. Molecule type: mRNA

A. Residues: 1-764 <RES>

A. Cross-references: EMBL: X7079; NID:9456345; EMBL:CAA51532.1; PDB:9456346

A. Note: submitted to the EMBL/Genbank/DBJ databases by J. F. Piskurich, February 1994

R. Kraljci, P. J. Kraljci, R. J. Sandberg, M. O. J. Jansen, T. J. Brandtzaeg, P.

Biochem. Biophys. Res. Commun. 156, 783-789, 1989

A. Title: Molecular cloning of the human transmembrane secretory component (poly-Ig re

A. Reference number: A32263, MIMD:89149794, PMID:2920039

A. Accession: A32263

A. Molecule type: mRNA

A. Residues: 72-764 <RES>

A. Cross-references: EMBL: X4459; NID:9514365; EMBL:AAA0104.1; PDB:9514366

R. Fallgren-Greuter, E. Greuter, W. Bastian, A. Kraljci, H. D. Eilert, H. J. Zimmer

Biol. Chem. Hoppe-Seyler 372, 119-128, 1991

A. Title: Determination of the molecular structure of the human free secretory compone

A. Reference number: S13453; MIMD:91315750; PMID:1859628

A. Accession: S13453

A. Molecule type: protein

A. Residues: 478-488; 517-526, 543-545 <FAL>

A. Note: disulfide bonds for unbound and IgA bound forms

K. Eilert, H. J. Quentel, E. J. Wiedelund, M. J. Hillemeier, S. J. Decker, J. J. Weber, M. J. Hill

Biol. Chem. Hoppe-Seyler 372, 119-128, 1991

A. Title: Determination of the molecular structure of the human free secretory compone

A. Reference number: S13453; MIMD:91315750; PMID:1859628

A. Accession: S13453

**AtTitle:** The primary structure of the human free secretory component and the arrangement  
**A:Reference number:** A02112; MIM:8512881; PMID:6526384  
**A:Accession:** A02112  
**A:Molecule type:** protein  
**A:Residues:** 19-157; D', 158-207; D'E', 210-238; D'F', 243; N', 245; 249; Q', 249-261; Q', 263-270.  
**A:Note:** paper in German with English abstract  
**C:Comment:** As a 100k transmembrane receptor for polymyxin immunoglobulin, secretory component forms interchain disulfide bonds, undergoes proteolysis and transcytosis. Free secretory component is secreted.  
**A:Gene:** GDB:PIGB  
**A:Cross-references:** GDB:120290, OMIM:173880  
**A:Map position:** 191-1941  
**A:Introns:** 15/1, 130/1, 149/1, 490/1, 509/1, 629/2, 670/1, 714/1, 733/3  
**A:Note:** The first intron occurs before the initiator codon  
**C:Complex:** monomeric as a transmembrane receptor or free in mucosal secretions; heterodimers, hetero-22-mers composed of one chain of secretory component, one chain of immunoglobulin heavy chain, one chain of immunoglobulin light chain  
**C:Superfamily:** secretory component, immunoglobulin receptor, phosphoprotein; transcytosis; duplication, glycoprotein, immunoglobulin receptor, phosphoprotein; transcytosis  
**Keywords:** duplication, glycoprotein, immunoglobulin receptor, phosphoprotein; transcytosis  
**F:1-18/Domain:** signal sequence \*status predicted <SIG>  
**F:19-764/Product:** transmembrane secretory component \*status predicted <MATF>  
**F:19-577/Product:** free secretory component \*status experimental <MATF>  
**F:33-112/Domain:** immunoglobulin homology <IM1>  
**F:145-222/Domain:** immunoglobulin homology <IM2>  
**F:250-322/Domain:** immunoglobulin homology <IM3>  
**F:364-443/Domain:** immunoglobulin homology <IM4>  
**F:475-546/Domain:** immunoglobulin homology <IM5>  
**F:619-661/Domain:** transmembrane \*status predicted <TM>  
**F:662-764/Domain:** intracellular \*status predicted <INT>  
**F:10-110-56-64-152-220-257-271-279-371-441-385-395-482-544-496-503/Disulfide bonds:** F:83-90, 135-186, 421-469, 499/Binding site: carbohydrate (Asn) (covalent) \*status experimental  
**F:486-520/Disulfide bonds:** (in 19-unbound form) \*status experimental  
**F:486-520/Disulfide bonds:** interchain (to 19a alpha-1 chain-192) \*status experimental  
**F:520/Binding site:** cysteine (Cys) (covalent) (in 19-bound form) \*status experimental  
**F:577-578/Cleavage site:** Lys-Ala (unidentified proteolysis) \*status experimental  
**F:673/Binding site:** phosphate (Ser) (covalent) \*status predicted

**Query Match** 8.9% Score 182; DB 1; Length 764;  
**Best Local Similarity** 37.4% Pred. No. 1.8e-05;  
**Matches** 43; Conservative 20; Mismatches 40; Indels 12; Gaps 5;

**QY** 12 PVSGARILPEVKEGSELGSGVTKKPLP---EMHVRILCEPMASGSGTGVTSTNF 67  
 DB 21 PIFG-----PE-FVNSVEGNSVSLICYPPTSVNHRHTKKYMGCGARCG-CITLISGCGY 73

**QY** 68 IKAERYGPTLKOYPPKNIPLVEVQLTESGVTAGAGMNTDGRKQKVLNV 122  
 DB 74 VSKVAGRANLTFENGTFVNVIALSDSGRYKCGIGINS-RGLSEFDSLEY 127

**RESULT 5**  
**ORRBG**  
**secretory component precursor - rabbit**  
**N:Alternate names:** poly-1a receptor; polymyxin immunoglobulin receptor  
**C:Comment:** free secretory component, transmembrane secretory component  
**C:Species:** Oryctolagus cuniculus (domestic rabbit)  
**C:Date:** 15-Nov-1984 \*sequence-revision 15-Nov-1984 \*text-change 16-Jul 1999  
**C:Accession:** A02111; A28077  
**R:Author:** K.E. Friedlander, M. Blobel, G  
**Nature** 308, 37-43, 1984  
**AtTitle:** The receptor for transferrin: transport of 19a and 19m contains multiple imm  
**A:Reference number:** A02111; MIM:8414246; PMID:6322002  
**A:Accession:** A02111  
**A:Molecule type:** mRNA  
**A:Residues:** 1-773 <MNS>  
**A:Cross-references:** GDB:X00413; GDB:K01701; NID:q1595; ETCN:CA25118.1; PID:q1596  
**A:Note:** The authors translated the codon ACC for residue 54 as Asn  
**J. Biol. Chem.** 263, 8120-8125, 1988  
**AtTitle:** Rabbit secretory components of different allotypes vary in their carbohydrate  
**A:Reference number:** A28077; MIM:80228032; PMID:3131393  
**A:Accession:** A28077  
**A:Molecule type:** protein  
**A:Residues:** 87-114;410-424 <PRU>

**C:Comment:** This receptor binds polymyxin 19a and 19m at the basolateral surface of cp  
**C:process:** cleavage occurs to separate the extracellular portion, also known as the sec  
**C:Comment:** The five domains exhibit homology with immunoglobulin V regions. The simi  
**C:Complex:** Alternative splicing in the extracellular domain leads to high or low mole  
**C:Superfamily:** secretory component, immunoglobulin homology  
**Keywords:** alternative splicing, duplication, glycoprotein; immunoglobulin receptor;  
**F:1-18/Domain:** signal sequence \*status predicted <SIG>  
**F:19-773/Product:** transmembrane secretory component \*status predicted <MATF>  
**F:19-577/Product:** free secretory component \*status predicted <MATF>  
**F:30-647/Domain:** extracellular \*status predicted <EXT>  
**F:39-117/Domain:** immunoglobulin homology <IM1>  
**F:148-227/Domain:** immunoglobulin homology <IM2>  
**F:253-326/Domain:** immunoglobulin homology <IM3>  
**F:362-443/Domain:** immunoglobulin homology <IM4>  
**F:471-540/Domain:** immunoglobulin homology <IM5>  
**F:448-670/Domain:** transmembrane \*status predicted <TM>  
**F:671-773/Domain:** intracellular \*status predicted <INT>  
**F:46-115-155-225-250-324-369-438-478-538/Disulfide bonds:** \*status experimental  
**F:108/Binding site:** carbohydrate (Asn) (covalent) (partial) \*status experimental  
**F:418/Binding site:** carbohydrate (Asn) (covalent) \*status experimental

**Query Match** 8.1% Score 166.5; DB 1; Length 773;  
**Best Local Similarity** 42.7% Pred. No. 9.0e-24;  
**Matches** 44; Conservative 11; Mismatches 41; Indels 7; Gaps 4;

**QY** 30 GGSVILKCP---EMHVRILCEPMASGSGTGVTSTNF 65  
 DB 39 GDSVITCYPTTSVTRHFKFEMPE-ESGRVILAS-IVSYSEYSGKRTLPDRG 96

**QY** 86 LPLVEVQLTESGVTAGAGMNTDGRKQKVLNV 128  
 DB 97 EFVVIWQILQNSDSYKCGVGN-QKGLDHPGVNLVLSQKPRP 138

**RESULT 6**  
**THHGL**  
**transforming protein gli - human**  
**N:Alternate names:** glioma-associated transforming protein  
**C:Species:** Homo sapiens (man)  
**C:Date:** 31-Mar-1991 \*sequence-revision 31-Mar-1991 \*text-change 10-Sep-1999  
**C:Accession:** S00672  
**R:Author:** K.W. Rapp, J.M. Bigner, S.H. Vogelstein, B.  
**Nature** 332, 371-374, 1988  
**AtTitle:** The GLI gene is a member of the Kruppel family of zinc finger proteins.  
**A:Reference number:** S00672; MIM:8815051; PMID:2832761  
**A:Accession:** S00672  
**A:Molecule type:** mRNA  
**A:Residues:** 1-1106 <KIN>  
**A:Cross-references:** EMBL:X07364; NID:q1767; ETCN:CA40247.1; PID:q1768  
**C:Gene:** GDB:GLI  
**A:Gene:** GDB:GLI  
**A:Cross-references:** GDB:119088; OMIM:162220  
**A:Map position:** 12q13.2-12q13.3  
**C:Superfamily:** gli transforming protein  
**C:Keywords:** DNA binding, duplication, oncogene; transcription regulation; transformin  
**F:225-260/Region:** zinc finger CCHH motif  
**F:268-295/Region:** zinc finger CCHH motif  
**F:301-325/Region:** zinc finger CCHH motif  
**F:331-356/Region:** zinc finger CCHH motif  
**F:362-387/Region:** zinc finger CCHH motif

**Query Match** 5.7% Score 117; DB 1; Length 1106;  
**Best Local Similarity** 22.3% Pred. No. 1.4;  
**Matches** 82; Conservative 33; Mismatches 130; Indels 122; Gaps 17;

**QY** 110 TDDCKTQKVLINHSYEPISWEFQPPPKMFIPLPOMAYASSSKFVIRVTPAQ- 168  
 DB 346 SUAKKQNK-----HSNKA-----FYVKKLQKQIK-----RTIDPSSL 379

**QY** 169 FGVPPVPHNSPTQTTP-----PVSRASSVAGKPPPTLPSTASTKISALEDLK 221  
 DB 380 RKHAKTIVH--GPDHVTKHKRQDPIPRAPASTVHPKREKREGCP1RDESKLIVPEGAMK 417

QY 222 LGPPSYNHTRL - HROKALVCSQSCREDCG ..... 251  
 DB 438 PQ-PSGAGSSSSSSGHSFPAASANTSSVEYMENAGSSTELLSDLEPCJAGTSLTLK 496  
 QY 252 ..... PHILIFT-ILGLEL - LALLGLVKAVERPVALSR-----ARPLAVP 292  
 DB 497 KLEMLIDOLHOIRPGTGLKLPSTLSHGTCTVSKRVGPPSLERSSSSSSISSAYVVS 556  
 QY 293 MPALESSQPPRSPFPFSSCNITVSAF-----PAFSAIAATIEAPVGP-----GAP 341  
 DB 557 RRSLSLSPPPGSPSPENASSLGLMPAPGHYLLPAPYASAPRGSTSPTAASSLIRIGSLP 616  
 QY 342 LPP-----APVQVESPMILHAPSLIKTSCRYVSLYHQPAAMHDS 380  
 DB 617 MPWSPRAEYPCVYNPAAVTPPASPAPAAADRP-----APAPVQFKSTGCVHTPTVAGG 673  
 QY 381 DS-DOYI 386  
 DB 674 QNDPPI 680

## RESULT 7

EDBERG

Immediate-early protein IE175 - Marek's disease virus (strain GA)  
 N:Alternate names: Infected cell protein 4  
 C:Species: Marek's disease virus  
 C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 16-Jul-1999  
 C:Accession: A42991  
 R:Anderson, A.S.; Francesconi, A.; Morgan, R.W.  
 Virology 189, 657-667, 1992  
 A:Title: Complete nucleotide sequence of the Marek's disease virus (CP4 gene, A:reference number: A42991, M0103251564, PMID:1322594  
 A:Accession: A42991  
 A:Molecule type: DNA  
 A:Residues: 1-1415 <AND>  
 A:Cross-references: GR M75729, NID 933650, P10N AAA46111, P10-g30951  
 C:Superfamily: herpesvirus immediate-early protein IE175  
 C:Keywords: DNA binding; early protein; transcription regulation

## Query Match

Best Local Similarity 5.5%, Score 113.5, DB 1, Length 1415;  
 Probable multi-domain beta keto-acyl synthase - Streptomyces coelicolor  
 Matches 70; Conservative 37; Mismatches 108; Indels 117; Gaps 14;

QY 104 CGAGMTDQKTKVTLNLSHYEPS--WEQPMDEPKWHLLPYLQMPAYASSSKPYT 161  
 DB 44 CNAHMDDD GQPMFPPLPYEDMNNSTDMRSPRSGPK---KDCGLDPLDTSG---P 96  
 QY 162 RVTTPAOKGKVPVHHSSITTOITHTPR.....VS 191  
 DB 97 RLTPPSS-GMSRLPHTSSPPKSPRPKGPETSPSEHIIISPPRNNTTHHNVCHVS 155  
 QY 192 RASSVAGDKPFTLSTT- ASKISALEGLKFCUPFSINHTPLIFQALDTGSGS 246  
 DB 156 RSPSSSSSSSSSSSSSSSIVLSSPSSSRSPSPRRASSSRPKRCSNCGKSG 215  
 QY 247 PFCGCHIIIPHTICFLIALDLVYKAVPEPKA:SPRAPRLAYRMALSSQHPKSP 306  
 DB 216 PQSKG-----RKA-SPTTKLEDEDTLPOETANRRGG 247  
 QY 307 PPRSNNIYSACPRAPCA-----IAACT--GFANVPGCAPLPFA 345  
 DB 248 RTRGRP- PKSGRAVGRND:QVTSSSGLADISPYDIAGSVWVEVILPPG----- 295  
 QY 346 FLQVSESPML-----IAPSLKTSCTFYVSLYHQ 372  
 DB 296 -----KCPWPGGLGCHPQALTPSPPLVPAVHR 321

## RESULT 8

CMRF-35 antigen - human  
 C:Species: Homo sapiens (man)  
 C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 21-Jul-2000

C:Accession: 137243  
 R:Jackson, D.G.; Hart, D.N.; Starling, G.; Bell, J.L.  
 Eur. J. Immunol. 22, 1157-1163, 1992  
 A:Title: Molecular cloning of a novel member of the immunoglobulin gene superfamily h  
 A:Reference number: 137243, M010322405, PMID:1349532  
 A:Accession: 137243  
 A:Status: preliminary; translated from GR/EMBL/JDBU  
 A:Molecule type: mRNA  
 A:Residues: 1-224 <RES>  
 A:Cross-references: EMBL X66171, NID 9306169, P10N AAA46148, P10-g306170  
 C:Genetics:  
 A:Gene: CMRF35

## Query Match

Best Local Similarity 5.5%, Score 112.5, DB 2, Length 224;  
 Probable multi-domain beta keto-acyl synthase - Streptomyces coelicolor  
 Matches 64; Conservative 33; Mismatches 100; Indels 87; Gaps 12;

QY 8 LYHLPVSGALRIIPVYKVGELGGSVTKCPRIEMH-VKTYLCHEMAGSCICGVSTT 65  
 DB 14 LILLIVPGYFPLSHDPTVAGPGSISVOCRYEKKRLTKFWGSP-POILIKCKIVETK 72  
 QY 66 NPIKAEYKSPVTLKQYPRKNLFLVEVTLLESISVAVVA:AMTERKTYKVLNVHSE 125  
 DB 73 G-SACKKNGKVSIRKSPANISFTVTLLENLDEDACTWCGV----- 112  
 QY 126 YEPSWEQPMPTPKW---PHLPYLQMPAYASSSKPYTRTPAOKOKVPVHHSSPTT 182  
 DB 113 -----DTP-WLRDFHDPIV-----EVEVSVFAGITTA-----SSP-- 142  
 QY 183 QTHRRPVRKASSVADIKRTLPSTTASKISALBGLIKQTPSYNHTTKIMKQALDYG 242  
 DB 143 -----USSMGTSGPPIKIVHWPSV-----KDSLPKSPH----- 174  
 QY 243 SUSGRGUGPHILPTLGL-PLALLSLVYKAVPEKALSR 285  
 DB 175 --VCSLFSWVKRLVLVLLPLLLSMIAVLMVNNVQVSSSKQ 216

## RESULT 9

137056

Probable multi-domain beta keto-acyl synthase - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
 C:Accession: T37056  
 R:Seeger, K.T.; Harris, D.; Parkhill, J.; Karelle, B.G.; Kalandream, M.A.  
 submitted to the EMBL Data Library, August 1999  
 A:Reference number: 221620  
 A:Accession: 137056  
 A:Status: preliminary; translated from GR/EMBL/JDBU

A:Molecule type: DNA

A:Residues: 1-2082 &lt;SEE&gt;

A:Cross-references: EMBL AL10747, P10N GAA43531, S5FER G300070, S5EDB S5221, 07

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOPDR:SC121.07

C:Superfamily: Streptomyces coelicolor probable polyketide synthase; 3-oxoacyl-acyl-  
 acyl carrier (protein) S-malonyltransferase homology  
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
 F148-479/domain: 3-oxoacyl-acyl-carrier-protein synthase I homology <ONS>  
 F1229/Binding site: phosphopantetheine (Ser) (covalent) \*status predicted

## Query Match

Best Local Similarity 5.5%, Score 112, DB 2, Length 2082;  
 Matches 73; Conservative 20; Mismatches 124; Indels 86; Gaps 12;

QY 105 GAGMNTDQKTKVTLNVHSEYPSWEQPMPTPKWHLLPYLQMPAYASSSKPYTV 163  
 DB 917 GACADTFHAALGELAVRCVLDLNLIMLPYASVPYPAKREPRKM-----TVK 963  
 QY 164 TTPADRGKVPV-VHHSSTTQTHPPVPSRASSVAGSKPPTFLPSTTASISALETLK 221  
 DB 964 ISCANYGQILPPSGTSHVEETTPADHGPPRPDAVPAVAAPVPAVALPSOAPGVV----- 1015

QY 222 PQTTSY-----NHHR-----LHQRALDYGSGSGREG 249  
 DB 1016 PLPEVFEAANSSPGEERSHPAEAPADAKVYGDDEQLAVESIHRTA-ETILALQRL 1074  
 QY 250 GCPHTLLPTILGLFLALLNIVKRAVEPRKALSRARPLAVPMALSSQPRGSPRRP 309  
 DB 1075 TDSHAPLKMTEFTLLALICAPVGAIVGMIA-----PAALPLPR 1114  
 QY 310 SONNI---YSACPPRRARADAAITGEA-PVPGRGAPILPPAPL---QVSE-SPLIAPSL 360  
 DB 1115 SVGLDPAVAAPVAAPVAAPVAAGATPYVPEASAPASAAKRPVSSEPAVPAQAPAR 1174  
 QY 361 KTS 363  
 DB 1175 DAS 1177

RESULT 10

Q0HE77  
glycoprotein 1 precursor - human herpesvirus 1 (strain 17)

C:Species: human herpesvirus 1

A:Note: host Homo sapiens (mas)

C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 16-Jul-1999

C:Accession: A05243

R:McGeoch, D.J., Dolan, A., Donald, S., Rixon, F.J.

J. Mol. Biol. 181, 1-13, 1985

A:Title: Sequence determination and genetic content of the short unique region in the gB

A:Reference number: A00656; M01D:55160822; PMID:2984429

A:Accession: A05243

A:Molecule type: DNA

A:Residues: 1-390 <MCC>

A:Cross-references: GB:L00046; NID:g291490; P1IN:AA96681.1; P1ID:g291497

C:Genetics:

A:Gene: US7

C:Superfamily: herpesvirus US7 protein

C:Keywords: glycoprotein, transmembrane protein

F:1-23/Domain. Signal sequence #status predicted <SIG>

F:24-390/Product. glycoprotein 1 #status predicted <GP1>

F:277-293/Domain. transmembrane #status predicted <TM>

F:156,175,257/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.48; Score 111.5; DB 1; Length 390;

Host local Similarity 23.28; Pred. No. 1.1; Indels 85; Gaps 16;

Matches 88; Conservative 35; Mismatches 171;

QY 17 LRLPEVVEGE-----EGSVTI-----KCPLEMHVRLYLQ--REMASSGTGT 60  
 DB 51 LLLIGELHFGVDQVPHITTYDGVHIMHYPMCHKCPVHVIVVTACPRRPAVAFALGRA 110  
 QY 61 VSTTNFKAKYKGVITIKOTPRKNLFLVETQLTSDSCVYA---CGAGMNTD----- 111  
 DB 111 TDSITIS--PAVPTLELNLAQOP-----LLRYQATRDYAGVYVLRVWVGAPNALSFLVLG 163  
 QY 112 KCKTDKVTLLNVHSEYDSEMEHOPETPRKMFHLPLQMPAVYSSSSGFVIRVITPA-ORG 170  
 DB 164 MAIAKCTLAANGSAYGCDIKILPSS-----AIRIAIAS VQIQAINDA 207  
 QY 171 KVRVHHSSPTTQITTHPRVSRASSVACDKPRFTLPSTTASKISALEGLKPOLTPSYNH 230  
 DB 208 STPTSTTISPTTIT-----PAISTTIPAPQASTTTPPGDKIKVDPVCVNH 253  
 QY 231 TKLHHQRALDYGSGSGREGPHILP-ILGLFLALLNIVKRAVEPRKALSRK----- 265  
 DB 254 PPSNATRTATROSRYALVTQITQIATPASTITAIIVLSGCTGTFRQ-QRPYPRSRPPIYS 312  
 QY 286 AAPLAVHRAIPSSQPRGSPRRPSONN---YSACPPRRARADAACT 330  
 DB 313 PQMPICISCAVNAFAMAMALQALIKSHDSTPRKSRHSSRTIMSLAIATSHRTAGAGL 372  
 QY 331 CEAPYPRGAPILPPAPLOY 349  
 DB 373 PTPPV-DPPTTPPPPLILV 390

RESULT 11

EB8108  
protein C46E10.3 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 18-May-2001 #sequence\_revision 18-May-2001 #text\_change 10-May-2001

C:Accession: EB8108

R:Anonymous, the C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A:Reference number: A75000; M01D:4006613; PMID:9451916

A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C

A:Accession: EB8108

A:Molecule type: DNA

A:Residues: 1-392 <STO>

A:Cross-references: GB:chr11; P1IN:AH96688.1; P1ID:g2773167; GSDH:GND00020; CESD:C46E

C:Genetics:

A:Gene: C46E10.3

A:Map position: 2

Query Match 5.48; Score 111.5; DB 2; Length 392;

Host local Similarity 23.78; Pred. No. 1.1; Indels 93; Gaps 20;

Matches 81; Conservative 47; Mismatches 121;

QY 88 LVEVTGLTESDQVYACANNTDQKTKVTLNVHSEYDSEMEHOPETPRKMF-HLPY 146  
 DB 9 LEEVTKLAEISIPDEL-----RGEPAQIVAVY--AREPSAE-PLPILPRVYKEEY 56  
 QY 147 LQMPAVYSSSSKPVTVKVTTPAQKCKVPVHHSSPTTQITTHPRVSRASSVACDKPRFTLP 206  
 DB 57 PFENSEQASSSP-----PVQILCASPNQGVIIH-----GKSLVAGRPDDRTIP 99  
 QY 207 STTAKISALELILKPVIT-PSYNNHTP-LHPQ---PALI-VYSGS-PPVQVFIILP--- 257  
 DB 100 ANTPRLVTPNIVTPANGPLQAPKLEKEDILKQIDVFRQSFQTAQNYQIVITDNK 159  
 QY 258 TLG-----LFLALLLSLVKRAVEPRKALSRARPLA---VPMALSS 269  
 DB 160 TLLSTQDLEQANTPTLSQVKSLSSENEAVK---EGDPATPRPQSLKENGPLEPELPS 216  
 QY 300 --QRPGRSPRRPSONNIVSACPRRARKADAGTGPANVPGCAP---LPP----- 344  
 DB 217 AVEWNGRPEPFPQCSVINFPISLAISS-----LPSTISPAITLPRIDNLAKVPSE 267  
 QY 345 APLVYS---ESPMLHAPSLKTSCEYVSLYHQFPAAMMEDSS 382  
 DB 268 AHPQISSLPISLPPLKSPSLDQDIKIACTKN---MQDKON 304

RESULT 12

T29018  
hypothetical protein ZK84.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000

C:Accession: T29018

R:Kirsten, J.

submitted to the EMBL Data Library, April 1995

A:Description: The sequence of C. elegans cosmid ZK84.

A:Reference number: Z20553

A:Accession: T29018

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-801 <KIR>

A:Cross-references: F001:023181; P1IN:AA-48234.1; ZSPR:GND00020; ZSPR:ZK84.1

C:Genetics:

A:Gene: ZK84.1

A:Map position: 2

A:Introns: 22/2; 45/3; 108/1

C:Superfamily: collagen alpha 1(1) chain, fibrillar collagen carboxyl-terminal homoto

Query Match 5.4%, Score 111, DB 2, Length 801.

Best Local Similarity 20.6%, Pred. No. 2.7;

Matches 72, Conservative 36, Mismatches 157, Indels 84, Gaps 12.

```

OY 30 GGSVTLKCHLPMHVRVLCREMAGSGTCIVSTTFNKAHYGRVTLKQYPRKNIPIV 89
DB 315 GMAATAPAPSSSEADAMAAPTDSANADTALVDTSS--SEHAAEST--EAPATDIAT 369
OY 90 EVYQVLTFSISDGYACGACNTDRCQVTLNVSHSYPRSMERDMP-----ETPRWPH 143
DB 370 EHTPAPSVAVIYVADAAGYDSSSTIVETP.....APAAETPATAASAARETPA--P 420
OY 144 LPYLQMPAYASSSKFVETVTPAQRKRVPPVHSSPTQTTHRPVSKASSVAQKPT 203
DB 421 AIAARETPATPTAS.....AAIYVACCAAVADVAATVATANTSSAQSAGSYD- 472
OY 204 FLPSTTASIKALHCLIKVITSYHHHKLKQKALDYQSSQKQCGTHLITLIGLR 263
DB 473 -VPSEPASEVTA-----PIVESAEAPSDSAPIG-PAASEPAPAPIEAP----- 515
OY 264 LIALGLVKKAVERRKALSRERRRIAVRMALISSQRPNSRQ-----NNIV 315
DB 516 -----ATDAATLTATAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 549
OY 316 SACPRRARRAD-----AMSTGAPVPQAPLPAPAPLOVSESPMLIAPS 359
DB 550 EETPAPAPADDETPAPAPAAETTPAPADAPAEETPAPADAPAEETPAPATA 598

```

RESULT 13

omega-conotoxin-sensitive N-type calcium channel alpha 1B-1 subunit (alternatively spliced)

C.Species: Homo sapiens (man)

C.Date: 04-Mar-1993 #sequence\_revision 18 Nov-1994 #text\_change 13-Sep-1998

C.Accession: A42566

R.Williams, M.E.; Brust, P.F.; Feldman, D.H.; Patliti, S.; Stelmason, S.; Maronfi, A.; Mol

Science 257, 389-395, 1992

A.Title: Structure and functional expression of an omega-conotoxin-sensitive human N-type

A.Reference number: A42566, MUID:9235886, PMID:1321501

A.Status: preliminary, not compared with conceptual translation

A.Molecule type: nucleic acid

A.Residues: 1-2339 -WIL-

A.Experimental source: IMR32, hippocampus

A.Note: sequence extracted from NCBI database (NCBI:109168)

C:Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match 5.4%, Score 119.5, DB 2, Length 2339,

Best Local Similarity 20.0%, Pred. No. 9.7;

Matches 82, Conservative 49, Mismatches 139, Indels 139, Gaps 18.

```

OY 75 KVTIKQYPRKNIPIVYQVLTFSISDGYACGACAMNTDRCQVTLNVSHSYPRSMERD 134
DB 1894 KATLEDTQVAVLKARVELKROKSSISLNSGALQNGESLKEV-----SWQTOR 1943
OY 135 MPEPRKWEFLP-----YLQMPAYASSS-----KEVTR-----VTTP 166
DB 1944 TQNAHRAHAPVTLRCHSHITIVQSCALAVYQVOMOSTIKRQVLPDLPOLDSOORAAM 2003
OY 167 AGRGVKVPVHNSP-----TTQITTPPVRSASVAGKPPPLPSTTASIKSALEGILKP 222
DB 2004 FLAAETQVPTLASMFKESISLACGPKRTHD-STTPDGF-----PFSQNS- 2048
OY 223 QRTSYVNHTRIKLR--QRALDYQSSQKQCGTHLITLIGLR----- 269
DB 2049 -SIHHNHHTHPDFPKRSLEK-TSLALMDG---AFS-...SAVGFILPREQ 2074
OY 270 -LVKFAVERRRKALSPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 318
DB 2095 PTCQKRRHPRHPRHPRHPRHPRHPRHPRHPRHPRHPRHPRHPRHPRHPRHPRH 2151
OY 319 PPRAR-----CADAAVTGAPVPT-----PAPAPAPAPAPAPAPAPAPAPAPAPAP 355

```

DB 2152 PTAQEPHPHGGSGVSNWSPLELSTASTPGRAPFETPTETPESTIYVANSII 2211

OY 356 HAPSLIKIS-----GVYSIYH-----YPAAMHNSISDRI 486

DB 2212 HPAQQTSLPAPSHCKLSKQ:SEHMAHLMQKHSQVATASCKISQDRI, 2260

RESULT 14

conserved hypothetical protein, *Delinococcus radiodurans* (strain R1)

C.Species: *Delinococcus radiodurans*

C.Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C.Accession: B75310

R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, P.J.

M.; Shui, M.; Yamahutan, J.; Lam, P.; Muldoon, D.; Utterback, L.; Zalewski, C.

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A>Title: genome sequence of the radioresistant bacterium *Delinococcus radiodurans* R1.

A.Reference number: A75250, MUID:20036896, PMID:10567266

A.Accession: B75310

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-528 <MHI>

A.Cross-references: AF004048, AB AF000514, NID 9444429, PION-AAF11681 1; PION-9645

A.Experimental source: strain R1

C.Genetics: DR2133

A:Map position: 1

Query Match 5.2%, Score 107.5, DB 2, Length 528,

Best Local Similarity 25.7%, Pred. No. 3;

Matches 61, Conservative 25, Mismatches 96, Indels 55, Gaps 11;

OY 149 QMAYVASSS-----KIVTKVTPAQRKRVPPVHSSPTQTTHRPVSKASSVAQKPT 203

DB 324 QVDAHPRSSARPRAPPRKPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPR 381

OY 204 -----FLPSTTASIKSALEGLKPLQTSYVNHHTKHLKQKALDYQSSQKQCGTHLITLIGLR 259

DB 382 PVTGPMPSATQGRISA-----PRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPR 428

OY 260 LDFLLALLGLVKKAVPR-----RKALSPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 311

DB 429 -----RAATRRSPRTSPPRAPASVAPSPPTTPAPANSAAAPNSPTPKTN 474

OY 312 NNIVYACPRKAKCAVAGTGAAPV--QCGAPLPAPAPLOVSESPMLIAPS 364

DB 475 WETPMPCHPSAASTSPSPFAPPPVHVS-FEPIA...FEFAM:PRHAPFKEN: 546

RESULT 15

hypothetical protein KIAA0324 - human (fragment)

C.Species: Homo sapiens (man)

C.Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 05-Nov-1999

C.Accession: T02345

C.Kricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Monk, C.; Saunders, B.; Robinson, D.

re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, D.

submitted to the EMBL Data Library, March 1998

A.Description: Sequencing of human chromosome 16p13.3.

A.Reference number: Z14664

A.Accession: T02345

A.Status: preliminary; translated from GR/EMBL/DBM

A.Molecule type: DNA

A.Residues: 1-1791 <RIC>

A.Cross-references: EMBL A7004443, NID 32900349, PION-AAF11681 1; PION-9645

C:Genetics: A:Map position: 16

A:Introns: 1610/2; 1706/2

A:Note: KIAA0324

Query Match 5.2%, Score 107, DB 2, Length 1791;

Best Local Similarity 23.5%, Pred. No. 13;



Matches 56; Conservative 30; Mismatches 100; Indels 52; Gaps 8;

```

UY 114 KTKVTLINHSEVEPMEQPMETPKMFLPYLFQMPAYASSKVTVTTPAQKGVP 173
Db 696 KSRFARRGSRSSPEPKTSPTPPRRSSRSPDLTKARLSRRSRAS--SSPETRSPTP 753
UY 174 PVHSSPTTQIIFRRPVSPAS--SVAGDKPETFLPSTIA----SKISALEGLIKQTPSY 227
Db 754 PHIRSRSPSVSSPEPKSRSSRRSRSSPRT--KTSRPGSRSPKPPGLOPSRSPSR 810
UY 228 NHITRLHQPALDYNSQ-----GPRGQFHILIFTILGLLALLGL 270
Db 811 RKRTRITRRRDNSSOSISRRRONSRSRKYTRRRRCQGYH----- 853
UY 271 VVKRAVERRRKALSPARPLAVPMRALESSQ--PPRGSPRP--PQNNITYSAVPPRAP 323
Db 854 --SRSPAQESPTSSPPRRGRSPPTSPKPSRSPSPAPMKRSRSPASPAITHRRSR 909

```

Search completed: November 24, 2002, 06:13:37  
 Job time : 27 secs



GenCode version 5.13  
Copyright (c) 1993 - 2002 Compugen Ltd

OM protein - protein search, using sw model1

Run on: November 24, 2002, 06:01:26 ; Search time 13 Seconds

(without alignments)  
1244,290 Million cell updates/sec

Title: us-09-651-150b-2

Perfect score: 2055

Sequence: 1 MDRMLPDLFPLVPSGALRL.....HOPAMMEDSDDDYINVA 390

Scoring table: HIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	195	9.5	771	PIGR_MOUSE	070570 mus musculi
2	185	9.0	769	PIGR_RAT	P15083 rattus norv
3	182	8.9	757	PIGR_BOVIN	P81265 bos taurus
4	182	8.9	764	PIGR_HUMAN	P01833 homo sapien
5	166.5	8.1	773	PIGR_KABIT	P01832 oryctolagus
6	121	5.9	102	PIGP_PIG	Q29244 sus scrofa
7	117	5.7	1106	GLI1_HUMAN	P08151 homo sapien
8	113.5	5.5	1415	ICP4_HSVMG	Q02363 marek's dis
9	112.5	5.5	224	CM35_HUMAN	Q08708 homo sapien
10	111.5	5.4	390	VELI_HSV1	P06487 herpes simp
11	109.5	5.3	2339	CCAB_HUMAN	Q00975 homo sapien
12	102	5.0	530	FWK2_MOUSE	P12507 mus musculi
13	101.5	4.9	1253	SHK2_HUMAN	Q09603 homo sapien
14	101.5	4.9	3530	MY15_HUMAN	Q06487 homo sapien
15	101	4.9	2439	CCAB_KABIT	Q05152 oryctolagus
16	100.5	4.9	876	AREA_EMENT	P17429 myricetella
17	100.5	4.9	1616	APX1_HUMAN	Q13796 homo sapien
18	100	4.9	384	DUS5_HUMAN	Q16690 homo sapien
19	100	4.9	741	GISE_MOUSE	Q08080 mus musculi
20	100	4.9	753	SK30_HUMAN	Q04993 homo sapien
21	99	4.8	430	CD19_CAVRO	P25917 cavia porce
22	98.5	4.8	116	STP2_PAT	P11101 rattus norv
23	98.5	4.8	341	PCG2_CAVRO	Q05113 cavia porce
24	98.5	4.8	495	TRP2_MOUSE	Q05144 mus musculi
25	98.5	4.8	847	CR22_HUMAN	P02273 homo sapien
26	98	4.8	554	CSE1_HUMAN	P09603 homo sapien
27	97	4.7	3649	ACVS_MOUSE	P27743 nocardia la
28	95.5	4.6	538	PVR2_HUMAN	Q02692 homo sapien
29	95.5	4.6	1833	ZPR2_HUMAN	Q02692 homo sapien
30	95	4.6	862	SM4D_HUMAN	Q02854 homo sapien
31	95	4.6	1042	SPAI_HUMAN	Q06f54 homo sapien
32	94.5	4.6	275	IAP_GWCP	P41436 cydia pomon
33	94.5	4.6	3149	TEGU_EBV	P03186 epstein-bar

## ALIGNMENTS

RESULT 1	PIGR_MOUSE	STANDARD:	PFT:	771 AA.
AC	070570:			
DT	30-MAY-2000 (Ref. 39, Created)			
DT	30-MAY-2000 (Ref. 39, Last sequence update)			
DT	15-JUN-2002 (Ref. 41, Last annotation update)			
DE	polymeric-immunoglobulin receptor precursor (Poly-Ig receptor) (PIGR)			
DE	[contains: secretory component].			
GN	PIGR.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6 X CHA; TISSUE=Liver;			
RX	MEDLINE=95138517; PubMed=7836758;			
RA	Piskunich T F, Blanchard M H, Youngman K R, France J A,			
RA	Kaelzel C S,;			
RT	*Molecular cloning of the mouse polymeric Ig receptor. Functional			
RT	regions of the molecule are conserved among five mammalian species.*;			
PL	J Immunol 154:1735-1747(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/SvEv;			
PX	MEDLINE=98072444; PubMed=9409786;			
RA	Martin M C, Collier E M, Lam J L, Li T W H, Wang J;			
RT	*Genomic cloning and structural analysis of the murine polymeric			
RT	receptor (PIgR) gene and promoter region.*;			
RL	gene 201:189-197(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129, TISSUE=Liver,			
RA	de Groot N, Vollebregt E, Lee S H, Verbeet M P, de Hoet H A,;			
RT	*Molecular cloning and exon-intron organization of the gene encoding			
RT	the murine polymeric immunoglobulin receptor.*;			
PL	Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.			
CC	- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IgM AND IgM AT THE			
CC	HAEMATOPOIETIC SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN			
CC	TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.			
CC	DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE			
CC	EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE			
CC	TRANSMEMBRANE SEGMENT (BY SIMILARITY).			
CC	1 SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.			
CC	- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.			
CC	- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.			
CC	-----			
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announcements">http://www.isb-sib.ch/announcements</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			

Q99943 homo sapien  
P44364 rattus norv  
P54258 rattus norv  
Q99944 homo sapien  
Q99948 rattus norv  
P53346 homo sapien  
P53344 saccharomyc  
P07354 gallus gall  
P26698 thodocorcus  
P48379 mus musculi  
Q42642 cochliodonu  
P17799 gallus gall

DR	EMBL:	U06431:	AAA67440.1:	-	
DR	EMBL:	U83434:	AAC53585.1:	-	
DR	EMBL:	U83427:	AAC53585.1:	JOINED.	
DR	EMBL:	U83428:	AAC53585.1:	JOINED.	
DR	EMBL:	U83429:	AAC53585.1:	JOINED.	
DR	EMBL:	U83430:	AAC53585.1:	JOINED.	
DR	EMBL:	U83431:	AAC53585.1:	JOINED.	
DR	EMBL:	U83432:	AAC53585.1:	JOINED.	
DR	EMBL:	U83433:	AAC53585.1:	JOINED.	
DR	EMBL:	Y16524:	CAA76272.1:	JOINED.	
DR	EMBL:	Y16525:	CAA76272.1:	JOINED.	
DR	EMBL:	Y16526:	CAA76272.1:	JOINED.	
DR	EMBL:	Y16527:	CAA76272.1:	JOINED.	
DR	EMBL:	Y16528:	CAA76272.1:	JOINED.	
DR	EMBL:	Y16529:	CAA76272.1:	JOINED.	
DR	EMBL:	Y16530:	CAA76272.1:	JOINED.	
DR	EMBL:	Y16531:	CAA76272.1:	JOINED.	
DR	EMBL:	Y16532:	CAA76272.1:	JOINED.	
DR	MGD:	MGI:103080:	P1gr.		
DR	InterPro:	IPR003599:	Iq.		
DR	InterPro:	IPR003006:	Iq_MHC.		
DR	InterPro:	IPR003600:	Iq_1like.		
DR	RefSeq:	PH00047:	igf. 5.		
DR	SMART:	SM00409:	Ig. 4.		
DR	SMART:	SM00410:	Ig_1like. 2.		
KW	Immunoglobulin domain; Repeat.	Transmembrane, Glycoprotein, Signal			
FT	SIGNAL:	1	78	POTENTIAL.	
FT	CHAIN	19	771	POTENTIAL.	
FT	CHAIN	19	611	POLYMERIC-IMMUNOGLOBULIN RECEPTOR.	
FT	DOMAIN	19	645	SECRETORY COMPONENT.	
FT	TRANSMEM	646	668	EXTRACELLULAR (POTENTIAL)	
FT	DOMAIN	669	771	POTENTIAL.	
FT	DOMAIN	33	117	CYTOPLASMIC (POTENTIAL).	
FT	DOMAIN	145	227	IG-LIKE V-TYPE DOMAIN 1.	
FT	DOMAIN	250	331	IG-LIKE V-TYPE DOMAIN 2.	
FT	DOMAIN	363	447	IG-LIKE V-TYPE DOMAIN 3.	
FT	DOMAIN	477	553	IG-LIKE V-TYPE DOMAIN 4.	
FT	DISULFID	40	110	IG-LIKE V-TYPE DOMAIN 5.	
FT	DISULFID	152	220	POTENTIAL.	
FT	DISULFID	257	324	POTENTIAL.	
FT	DISULFID	370	440	POTENTIAL.	
FT	DISULFID	484	546	POTENTIAL.	
FT	CARBOHYD	90	90	N-LINKED (GLYCNA	) (POTENTIAL)
FT	CARBOHYD	147	147	N-LINKED (GLYCNA	) (POTENTIAL)
FT	CARBOHYD	170	170	N-LINKED (GLYCNA	) (POTENTIAL)
FT	CARBOHYD	206	206	N-LINKED (GLYCNA	) (POTENTIAL)
FT	CARBOHYD	420	420	N-LINKED (GLYCNA	) (POTENTIAL)
FT	CARBOHYD	471	471	N-LINKED (GLYCNA	) (POTENTIAL)
FT	CONFLICT	159	159	A -> V (IN REF. 1).	
FT	CONFLICT	396	396	V -> A (IN REF. 1).	
FT	CONFLICT	620	620	C -> R (IN REF. 1).	
FT	SEQUENCE	771	AA:	84998	MM: 78c81402fc710730
					CMC04;

Query Match	9.5 %	Score 195	NR 1	Length 771
Best Local Similarity	34.5 %	Pred. No. 8.4e-07		
Matches	48	Conservative	21	Mismatches 54
				Indels 16
				Gaps 4

QY 10 FIVASALRIIDYKAEKMGCSYIKCPHPE-----MHVYLYGRMAAGSGTCGIVASTT 6  
Db 13 FGVSVKTRSPIDFGQEVSSITFSPVSIITPYVPEDTQSVNHHTRKYMCPQGA-SGMCTLLISSN 71  
QY 66 NFKAEYKSPVTLKQYQPEKNIETVEVTLQETESDSGYAVAGSNMGRGKTQKQVTLINVASE 12  
Db 72 GYLISKVYSGRANINNPENKNTFVNIHQIIVQDDIGSTKCLG--TSNKKGLSPDYSI,PLV--- 12  
QY 126 YEPVMEEQFMETPKMFIIL 144  
Db 128 -----SOVPELPSDTHV 139

```

RESULT 2
PGR_RAT
ID_PGR_RAT      STANDARD;      PRT;      769 AA

```

```

AC      PI5083; 01-APR-1990 (Rel. 14, Created)
DT      01-APR-1990 (Rel. 14, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Polymorphic immunoglobulin receptor precursor (Poly-IG receptor) (PIGR)
GN      PIGR.
OS      Rattus norvegicus (Rat).
OC      Eukaryota, Mammalia, Chordata, Craniata, Vertebrata, Eutelestomi;
OC      Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Rattus.
CA      NcbiTaxId=10116;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      TISSUE=Liver;
FX      MIMLINE-8937823; PubMed:2776882;
RA      Hanting C., Brakc B., Braghetta P., Iuzio J.P., Stancay K.K.;
RT      "Intracellular Targeting signals of polynetric immunoglobulin
RL      receptors are highly conserved between species."
RL      FEBS Lett. 254:177-183(1989).
CC      -1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
CC      BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
CC      TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
CC      DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
CC      EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE
CC      TRANSMEMBRANE SEGMENT.
CC      -1- SUBCELLULAR LOCALIZATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
CC      -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC      -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on way
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.sib.ch/announce/
CC      or send an email to license@sib.sib.ch)
CC      -----
DR      EMBL: X15741; CAA33758.1; -.
DR      PIR: S05407; S05407.
DR      InterPro: IPR003599; Ig.
DR      InterPro: IPR003006; Ig_MHC.
DR      InterPro: IPR003600; Ig_Like.
DR      Pfam: PF00047; Ig_5.
DR      SMART: SM00409; Ig_4.
DR      SMART: SM00410; Ig_Like; 2.
KW      Immunoglobulin domain, Repeat, Transmembrane, Glycoprotein, Signal.
FT      SIGNAL 1 18
FT      CHAIN 19 769 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
FT      CHAIN 19 611 SECRETORY COMPONENT.
FT      DOMAIN 19 643 EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 644 666 POTENTIAL.
FT      DOMAIN 667 769 CYTOPLASMIC (POTENTIAL).
FT      DOMAIN 33 117 IG-LIKE V-TYPE DOMAIN 1.
FT      DOMAIN 145 227 IG-LIKE V-TYPE DOMAIN 2.
FT      DOMAIN 250 331 IG-LIKE V-TYPE DOMAIN 3.
FT      DOMAIN 363 447 IG-LIKE V-TYPE DOMAIN 4.
FT      DOMAIN 477 553 IG-LIKE V-TYPE DOMAIN 5.
FT      DISULFID 40 110 POTENTIAL.
FT      DISULFID 152 220 POTENTIAL.
FT      DISULFID 257 324 POTENTIAL.
FT      DISULFID 370 440 POTENTIAL.
FT      DISULFID 484 546 POTENTIAL.
FT      CAPOHSD 50 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CAPOHSD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CAPOHSD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CAPOHSD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE 769 AA; 84794 MM; 5P44444400255A7 GP/64;
Query Match 9.0%; Score 185; DH 1; Length 769;
Best Local Similarity 34.5%; Pred. No. 4,56-06;
Matches 41; Conservative 21; Mismatches 41; Indels 16; Gaps 4;
30 GGSATTCGCLPE---MHVRIYLCDEMAAGSGGTGIVSTNFIKAQYKGGVTLIKYPRRN 85

```

CC	-1-	SUBCELLULAR LOCATION:	TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
CC	-1-	SIMILARITY:	BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC	-1-	SIMILARITY:	CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
CC	-		
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC		between the Swiss Institute of Bioinformatics and the EMBL outstation at	
CC		the European Bioinformatics Institute. There are no restrictions on its	
CC		use by non-profit institutions as long as its content is in no way	
CC		modified and this statement is not removed, usage by and for commercial	
CC		entities requires a license agreement (See <a href="http://www.isb.sib.ch/adnotice/c/">http://www.isb.sib.ch/adnotice/c/</a> )	
CC		or send an email to <a href="mailto:license@isb.sib.ch">license@isb.sib.ch</a> )	
CC			
DR	EMBL:	X15741:	CMA33758.1; -.
DR	PIR:	S05407:	S05407.
DR	InterPro:	IPR003599;	Ig.
DR	InterPro:	IPR003006;	Ig_MHC.
DR	InterPro:	IPR003600;	Ig_Like.
DR	Pfam:	PF00047;	Ig_5.
DR	SMART:	SM00409;	Ig_4.
DR	SMART:	SM00410;	Ig_Like; 2.
KM		Immunoglobulin domain, repeat, Transmembrane, Glycoprotein, Signal.	
FT	Signal	1	18
FT	CHAIN	19	769
FT	CHAIN	19	611
FT	DOMAIN	19	643
F1	TRANSMEM	b44	bbb
FT	DOMAIN	667	769
FT	DOMAIN	33	117
FT	DOMAIN	145	227
FT	DOMAIN	250	331
FT	DOMAIN	363	447
FT	DOMAIN	477	553
FT	DISEULEID	40	110
FT	DISEULEID	152	220
FT	DISEULEID	257	324
FT	DISEULEID	370	440
FT	DISEULEID	484	546
FT	CARBOHYD	90	90
FT	CARBOHYD	135	135
FT	CARBOHYD	206	206
FT	CARBOHYD	471	471
SV	SEQUENCE	769 AA:	84794 MW: 58444 kDa 40mg/55A7 Cpr/64:

```

Query Match      9.0%; Score 185; DH 1; Length 769;
Best local Similarity 34.5%; Pred. No. 4,56-06;
Matches 41; Conservative 21; Mismatches 41; Indels 16; Gaps 4;
30 GGSATTCPLPE---MHVRIYLCREMGAGSGTGVASTNTEIAAEYKGGVTLKQYPRRN 85

```

Db 33 GNSVITCYPTDTSVNRHTRKYMROGA-NRYCATLITSSWYLSKESVSASLINPENS 91  
 QY 86 LFLVVEVQVLTFSISGVYACCAQNMIDKQIKQVTLNHSLEYSWEQPMPTPKMHH 144  
 Db 92 TIVINAILTQEDTGSYKCGIG TIRNGLEFVSLEV.....SVEPFENDTHV 139

RESULT 4  
 PICK\_HOVIN STANDARD; PRT: 757 AA.  
 AC P81265;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Polymeric immunoglobulin receptor precursor (poly-Ig receptor) (PICK)  
 DE [Contains: Secretory component].  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NC NCBI\_taxonomy:9913;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 RC TISSUE=Small intestine, and Mammary gland;  
 RC MEDLINE=95186063; PubMed=788045.  
 RA Kuiseth M.A., Krajel P., Myklopost O., Rognes S.;  
 RT \*Cloning and characterization of two forms of bovine polymeric  
 RT immunoglobulin receptor cDNA.\*;  
 RL DNA 7-11 Bp. 14-251-256(1995)  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Mammary gland;  
 RC MEDLINE=96069604; PubMed=7590352;  
 RA Verbeet M.P., Vermeer H., Wierdman G.C., de Boer H.A., Lee S.H.;  
 RT \*Cloning and characterization of the bovine polymeric immunoglobulin  
 RT receptor-encoding cDNA.\*;  
 RL Gene 164:329-333(1995)  
 CC -1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE  
 CC NASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN  
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.  
 CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE  
 CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE  
 CC TRANSMEMBRANE SEGMENT.  
 CC -1- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN. ALSO SECRETED.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: FOUND IN MAMMARY GLAND, JEJUNUM, LUNG,  
 CC KIDNEY AND SMALL INTESTINE.  
 CC -1- PTM: IN THE ABSENCE OF DIMERIC IGA, SER-727 IS PHOSPHORYLATED  
 CC WHICH ALLOWS PICK TO FUNCTION NORMALLY.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 DR EMBL: L04797; AAC01620.1;  
 DR EMBL: X81371; CA57136.1;  
 DR InterPro: IPR001599; Iq  
 DR InterPro: IPR003006; Iq\_MHC  
 DR InterPro: IPR003600; Iq\_Like  
 DR Pfam: PF00047; Iq; 5  
 DR SMART: SM00409; Iq; 4  
 DR SMART: SM00410; Iq\_Like; 1  
 DR Immunoglobulin domain: Repeat; Transmembrane; Glycoprotein; Signal;  
 KW Polymorphism: Phosphorylation; Alternative splicing.

FT SIGNAL 1 18  
 FT CHAIN 19 757  
 FT CHAIN 19 599  
 FT CHAIN 19 632  
 FT TRANSMEM 633 653  
 FT DOMAIN 654 757  
 FT DOMAIN 33 117  
 FT DOMAIN 145 227  
 FT DOMAIN 250 331  
 FT DOMAIN 363 447  
 FT DOMAIN 474 550  
 FT DOMAIN 40 110  
 FT DISULFID 56 64  
 FT DISULFID 152 220  
 FT DISULFID 257 324  
 FT DISULFID 271 279  
 FT DISULFID 370 440  
 FT DISULFID 384 394  
 FT DISULFID 481 543  
 FT DISULFID 485 519  
 FT DISULFID 495 502  
 FT CARBOHYD 83 83  
 FT CARBOHYD 426 426  
 FT CARBOHYD 468 468  
 FT MOD\_RES 727 727  
 FT VARSPIC 129 346  
 FT VARIANT 29 29  
 FT VARIANT 142 142  
 FT VARIANT 404 404  
 FT VARIANT 413 413  
 FT VARIANT 435 435  
 FT SEQUENCE 757 AA; 8434 MW; 6434 MD; DEED7F0bAEnC6 CPO64;

Query Match 8 98; Score 182; DB 1; Length 757;  
 Host Local Similarity 34.18; Pred. No. 7.2e-06;  
 Matches 42; Conservative 21; Mismatches 44; Indels 16; Gaps 4;

QY 17 LFLVVEVQVLTFSISGVYACCAQNMIDKQIKQVTLNHSLEYSWEQPMPTPKMHH 62  
 Db 10 LFLVVEVQVLTFSISGVYACCAQNMIDKQIKQVTLNHSLEYSWEQPMPTPKMHH 68  
 QY 63 STNFTKREYGVPTLKRYPKNLFVEVTLQTFSSGVYVADGMNTQKQVTLNV 122  
 Db 69 SSEVSDVDVFRANITNPESTFVVDISHTKESYKCGIGSS-PGLNFEVSLV 127  
 QY 123 HSE 125  
 Db 128 SVD 130

RESULT 4  
 PICK\_HUMAN STANDARD; PRT: 764 AA.  
 AC P01833;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1986 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Polymeric immunoglobulin receptor precursor (poly-Ig receptor) (PICK)  
 DE [Contains: Secretory component].  
 GN PICK.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NC NCBI\_taxonomy:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=92039621; PubMed=1682231;  
 RA Krajel P., Grzeschik K.H., Gauris van Kessel A.H., Olaisen B.,  
 RA Brandzaeg P.;  
 RT \*The human transmembrane secretory component (poly-Ig receptor):  
 RT molecular cloning, restriction fragment length polymorphism and  
 RT chromosomal sublocalization.\*;  
 RL Hum. Genet. 87:642-648(1991).



CN PIGR.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 NX NCBI\_TaxID:9986;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:84142246; PubMed:6322802;  
 RA Mostov K.E., Friedlander M., Blobel G.,  
 RT "The receptor for transepithelial transport of IgA and IgM contains  
 RT multiple immunoglobulin like domains."  
 RL Nature 308:17-43(1984).  
 RN 12  
 RP SEQUENCE OF 87-114 AND 410-428.  
 RX MEDLINE:8828032; PubMed:313339.  
 RA Fritliker S., Hughes G.J., Hanly W.C., Jalon J.-C.;  
 RT "Rabbit secretory components of different allotypes vary in their  
 RT carbohydrate content and their sites of N-linked glycosylation."  
 RL J. Biol. Chem. 263:8120-8125(1988).  
 CC -1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE  
 CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN  
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.  
 CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE  
 CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE  
 CC TRANSMEMBRANE SEGMENT.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.  
 CC -1- POLYMORPHISM: THE SEQUENCE SHOWN IS THAT OF ALLOTYPES T62.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X00412; CAA25118.1; -;  
 DR PIR: A02111; QRRBG.  
 DR PIR: A28077; A28077.  
 DR InterPro: IPR003599; Iq\_1MC.  
 DR InterPro: IPR003600; Iq\_1like.  
 DR SMART: PF00047; Iq\_5.  
 DR SMART: SM00409; Iq\_5.  
 DR SMART: SM00410; Iq\_1like; 1.  
 KW Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal;  
 KW Polymorphism.  
 FT SIGNAL 1 18  
 FT CHAIN 19 773  
 FT CHAIN 19 615  
 FT DOMAIN 19 647  
 FT TRANSMM 648 670  
 FT DOMAIN 671 773  
 FT DOMAIN 39 122  
 FT DOMAIN 148 232  
 FT DOMAIN 253 331  
 FT DOMAIN 362 445  
 FT DOMAIN 471 545  
 FT DISULFID 46 115  
 FT DISULFID 155 225  
 FT DISULFID 260 324  
 FT DISULFID 369 438  
 FT DISULFID 478 538  
 FT CARBOHYD 88 88  
 FT CARBOHYD 108 108  
 FT CARBOHYD 418 418  
 FT VARIANT 88 88  
 FT VARIANT 94 94  
 FT VARIANT 101 108  
 FT VARIANT 110 110  
 S -> F (IN ALLOTYPES T63).  
 S -> F (IN ALLOTYPES T63).

SQ SEQUENCE 773 AA; 83886 MM; DP2C4AD2F1193C65 CR6C4;  
 Query Match 8.18; Score 166.5; DB 1; Length 773;  
 Best Local Similarity 42.78; Pred. No. 4; ye-05;  
 Matches 44; Conservative 11; Mismatches 41; Indels 7; Gaps 4;  
 QY 30 GSSVITKTP-----FMVRIYICPMAGSSTGIVGNTNPKAKPKGVTTKQYRKN 95  
 DB 39 GDSVITTYVPTTSVTRISRFKFWRE-ESGR-VTLA-TQVTSQSVGRKLTDEPKG 96  
 QY 86 LFLVEVTLTSDSVYAGAGMNTDRGKTQKVLNWHSEYEP 128  
 DB 97 EHVIVTDLIQMSGYSKGVGVN-GRGLDPCGVAVISQKRP 138  
 RESULT 6  
 PIGR\_PIG STANDARD; PRT; 102 AA.  
 AC 029244;  
 DT 30-MAY-2000 (Ref. 39, Created)  
 DT 30-MAY-2000 (Ref. 39, Last sequence update)  
 DT 30-MAY-2000 (Ref. 39, Last annotation update)  
 DE Polymeritic-immunoglobulin receptor (poly-Ig receptor) (PIGR)  
 DE (Fragment).  
 DE PIGR.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
 NX NCBI\_TaxID:9623;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Small intestine;  
 RX MEDLINE:96327607; PubMed-8672129.  
 RA Winterer A.K., Fretschel M., Davies W.;  
 RT "Evaluation and characterization of a porcine small intestine cDNA  
 RT library: analysis of 839 clones."  
 RL Mamm. Genome 7:509-517(1996).  
 CC -1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE  
 CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN  
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: F14851; CAA23294.1; -;  
 DR InterPro: IPR003600; Iq\_1like.  
 DR SMART: SM00410; Iq\_1like; 1.  
 KW Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein;  
 FT NON\_TER 1  
 FT NON\_TER 102  
 FT SEQUENCE 102 AA; 11405 MM; 842925.64B150B8 CR6C4;  
 Query Match 5.98; Score 121; DB 1; Length 102;  
 Best Local Similarity 34.78; Pred. No. 0; 018;  
 Matches 33; Conservative 21; Mismatches 31; Indels 10; Gaps 6;  
 QY 17 LRLI--DEKVEGELGSGVITKICPLP--MHRVYLREMAVSGTGIVASTNFKAE 71  
 DB 2 LQVAKPERHL-IYGLAKSVIFDCAISGMANVAKPLIC-QLNKQKTCNVVINTLKKRAVD 59  
 QY 72 YKGVTLKQYPRKNL-PLVEVTLTSDSVYAGAG 105  
 DB 60 FEGRIILT--FKENSHFSVHTIGLKEKDAHYLVG 92  
 RESULT 7

GLI1\_HUMAN STANDARD: PRT: 1106 AA.  
 ID GLI1\_HUMAN  
 AC P08151;  
 DT 01-AUG-1988 (rel. 08, Created)  
 DT 01-AUG-1988 (rel. 08, Last sequence update)  
 DT 15-JUN-2002 (rel. 41, Last annotation update)  
 DE Zinc finger protein GLI1 (glioma-associated oncogene) (oncogene GLI1)  
 GN GLI1 OR GLI1  
 OS Homo sapiens (Human);  
 OC Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE:88175051; PubMed:2832761;  
 RA Klinger K.W., Ruppert T.M., Rignier S.H., Vogelstein B.;  
 RT "The GLI gene is a member of the Kruppel family of zinc finger  
 RT proteins.";  
 RL Nature 332:371-374(1988).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 234-388.  
 RA MEDLINE:93391968; PubMed:8378770;  
 RA Pavletich N.P., Pabo G.O.;  
 RT "Crystal structure of a five-finger GLI-DNA complex: new perspectives  
 RT on zinc fingers.";  
 RL Science 261:1701-1707(1993).  
 CC -!- FUNCTION: MAY REGULATE THE TRANSCRIPTION OF SPECIFIC GENES DURING  
 CC NORMAL DEVELOPMENT. MAY PLAY A ROLE IN CRANIOFACIAL DEVELOPMENT  
 CC AND DIGITAL DEVELOPMENT, AS WELL AS DEVELOPMENT OF THE CENTRAL  
 CC NERVOUS SYSTEM AND GASTROINTESTINAL TRACT. IMPLICATED IN THE  
 CC TRANSCRIPTION OF SHH SIGNAL (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Nucleus.  
 CC -!- TISSUE SPECIFICITY: TESTIS, MYOMETRIUM, AND FALLOPIAN TUBE.  
 CC -!- INDUCTION: AMPLIFIED IN GLIOMASTOMA CELLS.  
 CC -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER  
 CC PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: X0784; CA30297.1;  
 DR PIR: S00672; YHUC1.  
 DR PDB: 2GLI; 23-JUL-97.  
 DR TRANSFAC: T00330;  
 DR Gene: HGN:4317; GLI.  
 DR MIM: 165220;  
 DR InterPro: IPR000822; Znf\_C2H2.  
 DR Pfam: PF00096; ZF-C2H2; 5.  
 DR PRINTS: PR00048; ZINC\_FINGER.  
 DR SMART: SM00355; ZNF\_C2H2; 5.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
 DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 5.  
 KW Zinc finger; Metal-binding; DNA-binding; Transcription regulation;  
 KW Nuclear protein; Repeat; Proto oncogene; 3D-structure.  
 FT DOMAIN 235 387 ZINC FINGERS.  
 FT ZN\_FING 235 260 C2H2 TYPE.  
 FT ZN\_FING 268 295 C2H2 TYPE.  
 FT ZN\_FING 301 325 C2H2-TYPE.  
 FT ZN\_FING 331 356 C2H2-TYPE.  
 FT ZN\_FING 362 387 C2H2-TYPE.  
 FT DOMAIN 1058 1055 ASP/GLU-RICH (ACIDIC).  
 SU SEQUENCE 1106 AA; 117904 MW; F29B1D1A6CD3D91 CRC64;  
 Query Match 5.7%; Score 117; DB 1; Length 1106;  
 Host Local Similarity 22.3%, Prod No. 0.6;  
 Matches 82; Conservative 33; Mismatches 130; Indels 122; Gaps 17,  
 QY 110 TORGTOKVTLNVIISEYFSEWEDPMPETPKWFLPYLFQMPAYASSKFTVITPAQ- 168

Db 346 SDPAKHQNT--HSNEK-----PYVTKLPYTK-----RTDESSL 379  
 QY 169 PKKVPVHHSSPTQTTHP-----PVSAASSAGIKKPTTHLPSTPASKISAMIGLIK 221  
 Db 380 RKHKIVH--GDAIVTKRHKGDLPRAPSTIVEKREKESGPIREESRLTVEKAMK 447  
 QY 222 PGTFSYNIHTPL HPQALDYSGSGSEPTG ..... 251  
 Db 436 PV-PSIQAASSSSSHPSAAGSAANTDSVEMIGNAQGSTHSLSDGPGIAGIGLSTLR 496  
 QY 252 ..... FILIFT ILGLFL TALLGVPRVAFEFALSF ..... AEFGLVR 292  
 Db 497 KLENLRDLQHLRLIGHGKLPSTLSHTGTIVSRKVPVSLERSSSSSSISAVTAS 556  
 QY 293 MAATSSQPPKSPPPSPNNIYSAPP-----PAPGADAGTGTAIVPQ-----GAP 341  
 Db 557 PRSLASPPSPSPSPENASSLPGIMPAQHLYLLAPVASARGGTSPTANSSLDRTIGLP 616  
 QY 342 LPP-----APLVSESPMLHAFSLKTSCEEYSLTHOPAMMEDS 380  
 Db 617 MPPWRSRAFPYGYNNACVTKRASIPQAALRP---AVAKYGRKSLQVHTVTVAGCG 673  
 QY 381 DS-DDYI 386  
 Db 674 QNFDPYL 680  
 RESULT 8  
 ID ICP4\_HSVMG STANDARD: PRT: 1415 AA.  
 AC 002362;  
 DT 01-OCT-1993 (rel. 27, Created)  
 DT 01-OCT-1993 (rel. 27, Last sequence update)  
 DT 01-FEB 1994 (rel. 28, Last annotation update)  
 DE Trans-acting transcriptional activator protein ICP4 (Immediate-early  
 DE protein IE175).  
 GN ICP4.  
 OS Marek's disease herpesvirus (strain GA) (MDHV).  
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Alphaherpesvirinae; Marek's disease-like viruses.  
 CC NCBI\_TaxID:10388;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE:92351564; PubMed:1322594;  
 RA Anderson A.S., Franconson A., Morgan R.W.;  
 RT "Complete nucleotide sequence of the Marek's disease virus ICP4  
 RT gene.";  
 RL Virology 189:657-667(1992).  
 CC -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE  
 CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING  
 CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS. IT IS  
 CC RECRUITED FOR THE SWITCH FROM IMMEDIATE-EARLY TO EARLY MODE OF  
 CC GENE EXPRESSION.  
 CC -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.  
 CC -!- FTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF  
 CC PHOSPHORYLATION.  
 CC -!- SIMILARITY: BELONGS TO THE ICP4/1E140/1E180 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: M75729; AAA46111.1;  
 DR PIR: AA2991; EDBEGA.  
 DR InterPro: IPR005205; Herpes\_ICP4\_C.  
 DR InterPro: IPR005206; Herpes\_ICP4\_N.  
 DR Pfam: PF03584; Herpes\_ICP4\_N; 1.  
 DR Pfam: PF03585; Herpes\_ICP4\_C; 1.



KW Early Protein: Transcription regulation; Trans-acting factor;  
 KM DNA-binding: Phosphorylation, Nuclear Protein.  
 FT DOMAIN 155 200 SER/PNO-RICH  
 SQ SOURCE: 1415 AA; 154936 MW; C08467BEEF4D1136 CRC64;  
 Query Match 5.5%, Score 113.5, DB 1; Length 1415;  
 Best Local Similarity 21.1%; Pred. No. 1.5;  
 Matches 70; Conservative 37; Mismatches 108; Indels 117; Gaps 14;

QY 104 CGACNNTKGTQVNTLVNHSHTPS--WELQMPETPKFHLVYLRQMAVSSSKVY 161  
 DB 44 GNANDDDD-VQPMPEPLPVPQMNPNSTDWSPSPSPK---KLPQSDLPAPITSG---P 96  
 QY 162 PFTTPAGPGKYPVHHSPTQTTHPP-----VS 101  
 DB 97 KLTTPSS-GKMSLPHITSSSPSPKRCPTSPSNHIIISPPRNPNSTTHRNQHS 155  
 QY 192 RASSVA-IDKPTLPSTT---ASKISALEGILKPGT-SYNNHTPLHF-GLALVYSGS 246  
 DB 156 RSPSSSSSSSSSSSLVSSSSSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 215  
 QY 247 KRCQGHLLPTTIGETIALUDGVKRAVPRKALSPRAHLAVKRALESQRRPGSP 306  
 DB 216 PQSKG-----FKA-SPTKRLDEDEYLFQETANKRGUG 247  
 QY 307 RPRSNNTIYSCPRRARGA-----DAAGT--GEAPV-PGAPLEPA 345  
 DB 248 RPKCHP-----PKSCAVGNIDQVSSSGIADTSPPYICGSVWMEVPLPPG----- 295  
 QY 346 PLVYSESEFWL-----HAPSLKTSCEYVSLHQ 372  
 DB 296 -----RCWFGSLGSHRQALDSEIVEAIHR 321

RESULT 9  
 CM35\_HUMAN  
 ID CM35\_HUMAN STANDARD; PRT; 224 AA.  
 AC Q08708;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE CMRF35 antigen precursor (CMRF-35).  
 GN CMRF35 OR CMRF35A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92249405; PubMed-1349532;  
 RA Jackson D G., Hart D N.J., Starling G., Reel J.L.;  
 RT "Molecular cloning of a novel member of the immunoglobulin gene  
 superfamily homologous to the polymorphic immunoglobulin receptor";  
 RL Eur J Immunol 22:1157-1163(1992)  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-21442079; PubMed-11556966;  
 RA Clark G.J., Cooper H., Fitzpatrick S., Green H.J., Hart D N.J.;  
 RT "The gene encoding the immunoregulatory signaling molecule CMRF-35A  
 localized to human chromosome 17 in close proximity to other members  
 of the CMRF-35 family";  
 RL Tissue Antigens 57:415-423(2001)  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-1lung;  
 RA Strausberg R.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC -1- TISSUE SPECIFICITY: PRESENT ON THE SURFACE OF MONOCYTES,  
 CC NEUTROPHILS, A PROPORTION OF PERIPHERAL BLOOD T AND B LYMPHOCYTES  
 CC AND LYMPHOCTIC CELL LINES.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL databank -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X66171; CAA46948.1;  
 DR EMBL: AF373865; AAK64272.1;  
 DR EMBL: AF373863; AAK64272.1; JOINED.  
 DR EMBL: AF373864; AAK64272.1; JOINED.  
 DR EMBL: AF373865; AAK64272.1; JOINED.  
 DR EMBL: BC022279; AAK22279.1;  
 DR MIM: 606786;  
 DR InterPro: IPR003599; 1g.  
 DR InterPro: IPR003006; 1g\_MUC.  
 DR Pfam: PF00047; 1g; 1.  
 DR SMART: SM00409; 1g; 1.  
 KW Antigen; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 224  
 FT DOMAIN 21 183  
 FT TRANSMEM 184 204  
 FT DOMAIN 205 224  
 FT DOMAIN 36 117  
 FT DOMAIN 123 175  
 FT DISULFID 43 110  
 FT DISULFID 57 65  
 FT CARBOHYD 90 90  
 FT CARBOHYD 99 99  
 SQ SEQUENCE: 224 AA; 24830 MW; 60C8B716D84600D2 CRC64;

Query Match 5.5%, Score 112.5, DB 1; Length 224;  
 Best Local Similarity 22.5%; Pred. No. 0.19;  
 Matches 64; Conservative 33; Mismatches 100; Indels 87; Gaps 12;

QY 8 LYLIVASGALRIPLPEVKEGICGVTIKCPILPMH--VRIYICREAGSGTCGVSTI 65  
 DB 14 LLLILVPGYPLSHMTYAGPVGGSLSVQCYKPKHKLTKWKP-VLLIKCKIVETK 72  
 QY 66 NFKAEYGVATLKQYPRKNLFEVETQITFESGVYACGAMNTDQKTYTLNVHSE 125  
 DB 73 G-SACKRKQKVASIRSPANLISFTVLENIIDELAGIYQCV----- 112  
 QY 126 YRPSWEDQPMPTPKW---FHLPLYLQMPAYASSSKVTVRTTIAQCKVYVHHSP 182  
 DB 113 -----DTP-WLRDFHDPV-----EVEVSVEPAGTITA-----SSP-- 142  
 QY 183 QITHRPVRSASSVAGCPRTFTLSTASKISALEGILKPGT-SYNNHTPLHF-GLALV 242  
 DB 143 -----QSSMCTISDPIKILVHTWPSVT-----RKISHPSPSH----- 174  
 QY 243 SOSGREGGFHLIPTIGL-FLALGLGVKRAVEPKALSP 285  
 DB 175 --PGSIFSNVRLVLELPLILSMGAVLWVRPQSRSSK 216

RESULT 10  
 VGLI\_HSV1  
 ID VGLI\_HSV1 STANDARD; PRT; 340 AA.  
 AC P06487;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Glycoprotein I.  
 GN GI OR US7.  
 OS Herpes simplex virus (type 1 / strain 17).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10299;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA MEDLINE-85160822; PubMed 2984129;  
 RX McGeeoch D.T., Dolan A., Donald S., Nixon F.T.  
 RT "Sequence determination and genetic control of the spot unique  
 RI region in the genome of herpes simplex virus type 1.";  
 RL J. Mol. Biol. 181:1-13(1985).  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND  
 CC -1- SIMILARITY TO OTHER HERPESVIRUSES GLYCOPROTEINS 1, TO VZV GP1, AND TO PRV GP63.  
 CC AND TO PRV GP63.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1st-sib.ch/announce/>  
 CC or send an email to [license@sib-sb.ch](mailto:license@sib-sb.ch)).  
 CC -----  
 DR EMBL: L00036; AAA96681.1;  
 DR EMBL: X14112; CAJ2284.1;  
 DR EMBL: X02138; CAJ26061.1;  
 DR PIR: A05243; Q0BE77.  
 DR InterPro: IPR002874; Herpes\_gf.  
 DR Pfam: PF01688; Herpes\_gf\_1.  
 DR GlycoProfile:  
 FT CAROHD 156 156 N-LINKED (GLCNAC...) (POTENTIAL)  
 FT CAROHD 175 175 N-LINKED (GLCNAC...) (POTENTIAL)  
 FT CAROHD 257 257 N-LINKED (GLCNAC...) (POTENTIAL)  
 SQ SROHNC 390 AA: 41369 MW: 39381HIDB5F08C CNGC4;  
 Query Match 5.4%; Score 111.5; DI 1; Length 390;  
 Best Local Similarity 23.2%; Ered. No. 0.44; Matches 85; Gaps 16;  
 Matches 88; Conservative 35; Mismatches 171; Indels 85; Gaps 16;  
 QY 17 LRLPEVKEGE-----LGSVTI-----KCPLEPMHVRILYC--REMASTGCT 60  
 DB 51 LLLGLRLFRVGDQVPHPTTYDGVLMHYPMCKCPRVAVVTVACPRPAVAFALORA 110  
 QY 61 VVSTFTIAEYKRGVTLKQYPRKMLFVETQLTESQYVA--CGAGMND----- 111  
 DB 111 TDSIHS--DAYPTLEINLAQGP-----LKKOKARIDYAGVYVLRVWYSLANSLIPLVG 163  
 QY 112 KCKTVKIVLVNHSYPMSEWQPMPTPKMFIPLYLPMYVASSSKFVTVTTPA--ORG 170  
 DB 164 MALMAGTLAYMSAVGSDPKLIPSS-----APPLAAPS-----VYGPAPRQA 207  
 QY 171 KVPVHSSPTTQITRRVRSASSVADCKRTPLPSTTASISALKQLKPPPSYNNH 230  
 DB 208 STSTSTSTSTSTI-----FAVSTTIPAVQASTTTPPTGDPKDPGCVNHE 253  
 QY 231 TRLHPKALVYSGSGFETFHIIIP--TIIQLFLALL--LVKKAVFRPALSEF.... 285  
 DB 254 PPSNATRAIRDSRYALIVQIIQIAIPASTIALVFLGSCIFTHGC--QRRKRNSRPDYS 312  
 QY 286 ..ARRIAYNMALPSSOPRCPSPRSQNN-----YSACPRRANGADACT 330  
 DB 313 PMPPTCTCAVNEAMARAGALKSHIPSTPKSRKRSKRTMPSLTALAESEVAMAGL 372  
 QY 331 GIAPVPGCAPLPPIAPVQY 349  
 DB 373 PTPPV-DPTTPPTPIPLV 390  
 RESULT 11  
 CCAH\_HUMAN STANDARD; PRT; 2339 AA.  
 AC 000975;  
 DI 01-OCT-1996 (Rel. 34, Created)  
 DI 01-OCT-1996 (Rel. 34, Last sequence update)  
 DI 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Voltage-dependent N-type calcium channel alpha-1b subunit (Calcium  
 channel, L-type, alpha-1 polypeptide isoform 5) (Brain calcium channel)

DE 111 (H111).  
 GN CACNA1B OR CACNA1A5 OR CACNA5.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC NCBI\_Taxid:9606;  
 RN 11  
 RP SEQUENCE FROM N.A. (ISOFORMS ALPHA-1B-1 AND ALPHA-1B-2).  
 KC TISSUE-Brain;  
 KC TISSUE-Brain;  
 RX MEDLINE-92335886; PubMed-1321501;  
 RA Williams M.E., Brust P.F., Feldman D.H., Palith S., Simeron S.,  
 RA Maroufi A., McQue A.F., Veliceladi G., Ellis S.H., Harpold M.M.;  
 RT "Structure and functional expression of an omega-conotoxin-sensitive  
 RL human N-type calcium channel.";  
 RL Science 257:389-395(1992).  
 RN 12  
 RP SEQUENCE OF 1-94 FROM N.A.  
 KC TISSUE-lung fibroblast;  
 RA Kim D.S., Jung H.H., Park S.H., Chin H.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDATE THE  
 CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED  
 CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE  
 CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,  
 CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1B  
 CC GIVES RISE TO N-TYPE CALCIUM CURRENTS. N-TYPE CALCIUM CHANNELS  
 CC BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND ARE BLOCKED  
 CC BY (omega)-CONOTOXIN-IVA (omega-CTX-IVA) AND BY (omega)-AGATOXIN-  
 CC IIA (omega-AGA-IIA). THEY ARE HOWEVER INSENSITIVE TO  
 CC DIHYDROPYRIDINES (DHP), AND (omega)-AGATOXIN-IVA (omega-AGA-IVA).  
 CC CALCIUM CHANNELS CONTAINING ALPHA-1B SUBUNIT MAY PLAY A ROLE IN  
 CC DIRECTED MIGRATION OF IMMATURE NEURONS.  
 CC -1- STRUCTURE: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT  
 CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS  
 CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-  
 CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS  
 CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM  
 CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA  
 CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: ALPHA-1B-1 (SHOWN HERE) AND  
 CC ALPHA-1B-2; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: ALPHA-1B-1 AND ALPHA-1B-2 ARE EXPRESSED IN THE  
 CC CENTRAL NERVOUS SYSTEM, BUT NOT IN SKELETAL MUSCLE OR AORTA.  
 CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE  
 CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE  
 CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A  
 CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.  
 CC -1- PFM: PHOSPHORYLATED IN VITRO BY CAM-KINASE II, CAPK, PKC AND GSK3  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS  
 CC FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1st-sib.ch/announce/>  
 CC or send an email to [license@sib-sb.ch](mailto:license@sib-sb.ch)).  
 CC -----  
 DR EMBL: M94172; AAA51897.1;  
 DR EMBL: M94173; AAA51898.1;  
 DR EMBL: U76666; AAC51138.1;  
 DR Genew: HGNC:1389; CACNA1B.  
 DR MIM: 601012;  
 DR InterPro: IPR001682; Ca/Na\_pore.  
 DR InterPro: IPR002077; Ca\_channel.  
 DR InterPro: IPR003111; Cal\_channel\_Tfpl.  
 DR InterPro: IPR000636; Mechanel\_nlg.  
 DR Pfam: PF00520; ion\_Trans\_4.  
 DR PRINTS: PR00167; CACNANFL.



DR	EMBL:	M80206:	AAA39734.1		
DR	EMBL:	D26107:	BAA05103.1		
DR	PIR:	A38211:	ILMSP3.		
DR	MGD:	MG1:97822:	PVS.		
DR	InterPro:	IPR003599:	Ig.		
DR	InterPro:	IPR003006:	Ig_MHC.		
DR	InterPro:	IPR003600:	Ig_Like.		
DR	InterPro:	IPR003596:	Ig_V.		
DR	Plam:	PF00047:	Ig_3		
DR	SMART:	SM00409:	IG_3.		
DR	SMART:	SM00410:	IC_Like. 1		
DR	SMART:	SM00406:	ICV_1		
KW	Immunoglobulin domain; Receptor; Transmembrane; Glycoprotein; Signal;				
KW	Repeat; Alternative splicing.				
FT	SIGNAL	1	31		POTENTIAL.
FT	CHAIN	32	530		POLIOVIRUS RECEPTOR RELATED PROTEIN 2.
FT	DOMAIN	32	351		EXTRACTETULAR (POTENTIAL).
FT	TRANSMEM	352	372		POTENTIAL.
FT	DOMAIN	373	530		CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	47	138		IG-LIKE V-TYPE DOMAIN.
FT	DOMAIN	167	236		IG-LIKE C2-TYPE DOMAIN 1.
FT	DOMAIN	267	327		IG-LIKE C2-TYPE DOMAIN 2.
FT	DI.SUPEID	54	131		BY SIMILARITY.
FT	DISUPEID	174	229		BY SIMILARITY.
FT	DISUPEID	274	320		BY SIMILARITY.
FT	CARBHYD	128	128		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBHYD	138	138		N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CAPROHYD	315	315		N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	VARSPLIC	339	467		FSPSTAGACATGGCCTGGAALATATAACTGCTGTCGROORKE
FT					ORIOAABREBLEEGPSISYKPPPKARIIEPHMPSPITLIGA
FT					SEHSVPKPYFLDAVSADQEMPVHEHPTLEERSPLLGS
FT					ATGCG -> DTGASHVA; PLVM; AV; N1LLVLLASGEL
FT					ALLILGRFRFKSRGSHNDIDPGSYTPGVESGSSVEVM
FT					RASLDIEMRDGDREDEEREEERKKAREGIMLPNHSXPDDM
FT					FSHDIGSLISRAAYV (IN ISOFORM ALPHA).
FT					MISSING (IN ISOFORM ALPHA).
FT	VARSPLIC	468	530		DEDVDIFA;ECSLIBE CRCQ4.
FT	SEQUENCE	530 AA:	57417 MW:		
SC					

```

QY      8  LYLPLPSNA-----LPLTPEYVAVDELLBSASTIGCP-LPEMHWIRYLCEMGASSTIGCTV  62
           |   |::|   |::|::|   |::|::|   |::|::|   |::|::|   |::|::|   |::|::|
QY      23  LLLLOETGAGDVRVAVLEEV-GRSGSYELPCHLPPPTERV---SQYIWRKLDGTV  77
           |   |::|   |::|::|   |::|::|   |::|::|   |::|::|   |::|::|   |::|::|
QY      63  SII-----INFLIKELY-KGVYT-LKQYPRKNILPLEVT  QLTISDSQYVACCAAM-- 108
           |   |::|   |::|::|   |::|::|   |::|::|   |::|::|   |::|::|   |::|::|

```

	RESULT 13
SIHK2_HUMAN	
ID	SIHK2_HUMAN
AC	Q9UPX8; Q9UPK1; STANDARD; PRT, 1253 AA.
UT	15-JUN-2002 (Rel. 41, Created)
VF	15-JUN-2002 (Rel. 41, Last sequence update)
VF	15-JUN-2002 (Rel. 41, Last annotation update)
CE	SH3 and multiple ankyrin repeat domains protein 2 (Shank2).
CN	SHANK2 OR KIAA1022.
OS	Homo Sapiens (Human).
OC	Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;
OC	Mammalia, Eutheria, Primates, Catarrhini, Hominidae, Homo.
OX	NCBI_Taxid=9606;
RN	[1]
RP	CONCEPTUAL TRANSLATION OF 1-11.
KA	Reichert B.;
KL	Unpublished observations (JUL-2002).
RE	[2]
RF	PARTIAL SEQUENCE FROM N.A. (ISOFORM 4);
RX	MEDLINE-99436166; PubMed-10506216;
RA	Lim S., Natsubiti S., Yoon T., Hyang T.I., Suh P.G., Sheng M., Kim E.;
KT	"Characterization of the Shank family of synaptic proteins. Multiple
KT	genes, alternative splicing, and differential expression in brain and
KT	development."?
PL	J Biol Chem. 274:47951e-47951k(1999).
PN	[3]
PP	SEQUENCE OF 123-1253 FROM N A (ISOFORM 1).
RC	TISSUE=Brain;
RX	MEDLINE-99397452; PubMed-10470851;
RA	Kikuno R., Nagase T., Ishikawa K., Hirosewa M., Miyajima N.,
RA	Tanaka A., Kotani H., Nemura N., Ohara O.;
FT	"Prediction of the coding sequences of unidentified human genes. XIV.
FT	The complete sequences of 106 new cDNA clones from brain which code
FT	for large proteins in vitro."
KL	DNA Res. 6:197-205(1999).
KN	[4]
FP	INTERACTION WITH DNM2.
FX	PubMed-11583995;
FA	Okamoto P.M., Gamby C., Wells D., Fallon J., Vallbo A.B.;
FT	"Dynamism Isoform-specific interaction with the Shank/proSAP
FT	scaffolding proteins of the postsynaptic density and actin

RT CYTOSKELETON.\*  
 RL J. Biol. Chem. 276:48458-48465(2001).  
 RP REVIEW.  
 RX PubMed:10806096;  
 RA Sheuq M., Kim E.;  
 RT \*The Shank family of scaffold proteins.\*  
 RL J Cell Sci 113:1851-1856(2000)  
 CC -1- FUNCTION: Seems to be an adaptor protein in the postsynaptic  
 CC density (PSD) of excitatory synapses that interconnects receptors  
 CC of the postsynaptic membrane including NMDA-type and metabotropic  
 CC glutamate receptors, and the actin-based cytoskeleton. May play a  
 CC role in the structural and functional organization of the  
 CC dendritic spine and synaptic junction.  
 CC -1- SUBUNIT: Interacts with DNM2. Interacts with COTN/contactin SH3  
 CC domain, DlgAP1/GKAP and alpha-latrotoxin receptor 1 (By  
 CC similarity). Is part of a complex with Dlg4/PSD-95 and DlgAP1/GKAP  
 CC (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of  
 CC neuronal cells (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS: At least 2 isoforms; 1 (shown here)  
 CC and 2; are produced by alternative splicing.  
 CC -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB028945; AAA82974.1;  
 DR EMBL: AF141901; AAA02496.1;  
 DR MIM: 603290;  
 DR InterPro: IPR001660; SAM  
 DR Pfam: PF00536; SAM; 1;  
 DR SMART: SM00454; SAM; 1;  
 DR PROSITE: PS0106; PDZ; 1;  
 DR PROSITE: PS0105; SAM DOMAIN; 1;  
 DR PROSITE: PSC002; SH3; 1;  
 KW SH3-binding; Coiled coil; Alternative splicing.  
 FT DOMAIN 37 131  
 FT SITE 727 744  
 FT SITE 948 954  
 FT DOMAIN 1190 1254  
 FT DOMAIN 299 307  
 FT DOMAIN 796 799  
 FT DOMAIN 982 990  
 FT VARSPLIC 173 173  
 FT VARSPLIC 255 273  
 FT VARSPLIC 274 1254  
 FT SEQUENCE 1254 AA: 134799 MW: 08D46578C3E26C21 CRC64.  
 Query March 4 98; Score 101 5; nr 1; length 1254;  
 Host Local Similarity 25.48; Pred. No. 9.3;  
 Matches 59; Conservative 34; Mismatches 88; Indels 51; Gaps 15.  
 UY 160 VTRPTTP-ADGKVPVHHSPTQITRRPVRSS--VAGKRPITLPSITASKI 213  
 DB 130 VTRKADVDYAKKAPPPKAPPTALILRSK-SMTSELELVDRKPEIVASPSR- 187  
 UY 214 VALEGL LKP .....QTPSYNHTLRRLDYGSS-QSGREGSGFILLITIGLEL 264  
 DB 188 -AAENMAVEPPVATIKGPPS-----SPTEPAGSDNNVSEGGIIVMTPTPGSKK 237  
 UY 265 IALLIGLVKAAVRKKAISRAKKAIVAKKALHSSQPGSPPPSSQNNVYACD--RPA 322  
 DB 238 APLFLG-IPRGTMPRRKPSIDSP-----TFSGITEERQ-----FLAFPLMKET 279

UY 323 KGNLAAGVGEAFVPGAPLPPAPLAVSESPWLHAPSLKISCEVSLYIUPA 374  
 DB 280 KSLSMPTSR-DIPPPQSGVPPSPPPSPPTTY-NCGRKSPVPRVYGI1--KPA 327  
 RESULT 14  
 ID MY15\_HUMAN STANDARD, FBI; 3530 AA.  
 AC 090KN7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Myosin XV (unconventional myosin-15).  
 GN MYO15A OR MYO15.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; (Euteleostomi);  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxId:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.  
 RX MEDLINE:4001762; PubMed:10552926;  
 RA Liang Y., Wang A., Belyantseva I.A., Anderson D.W., Probst F.J.,  
 RA Barber T.D., Miller W., Touchman J.W., Jin L., Sullivan S.L.,  
 RA Sellers J.R., Camper S.A., Lloyd R.V., Kachar B., Friedman T.B.,  
 RA Fridell R.A.;  
 RT \*Characterization of the human and mouse unconventional myosin XV  
 RT genes responsible for hereditary deafness DFNB3 and Shaker 2.\*;  
 RL Genomics 01:243-258(1999).  
 RN [2]  
 RP PARTIAL SEQUENCE FROM N.A. AND VARIANTS DFNB3 1YR-2111 AND PHE-2113.  
 RX MEDLINE:98267311; PubMed:9603736;  
 RA Wang A., Liang Y., Fridell R.A., Probst F.J., Wilcox E.R.,  
 RA Touchman J.W., Morton C.C., Morell R.J., Noben-Trauth K., Camper S.A.,  
 RA Friedman T.B.;  
 RT \*Association of unconventional myosin MYO15 mutations with human  
 RT nonsyndromic deafness DFNB3.\*;  
 RL Science 280:1447-1451(1998).  
 CC -1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE  
 CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.  
 CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS  
 CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY  
 CC SIMILARITY). MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF  
 CC THE ACTIN-RICH STRUCTURES OF THE INNER EAR SENSORY HAIR CELLS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PITUITARY, ALSO EXPRESSED  
 CC AT LOWER LEVELS IN ADULT BRAIN, KIDNEY, LIVER, LUNG, PANCREAS,  
 CC PLACENTA AND SKELETAL MUSCLE. NOT EXPRESSED IN BRAIN. IN THE  
 CC PITUITARY, HIGHLY EXPRESSED IN ANTERIOR GLAND CELLS.  
 CC -1- DISEASE: Defects in MYO15A are the cause of an autosomal recessive  
 CC form of nonsyndromic deafness (DFNB3).  
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 HAND 4.1-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 10 DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF144094; AAF05903.1;  
 DR EMBL: AF051976; -; NOT\_ANNOTATED\_GUS.  
 DR HSSP: P10587; IBR2.  
 DR Genew: HGNC:7594; MYO15A.  
 DR MIM: 602666;  
 DR MIM: 600316;  
 DR InterPro: IPR000299; Hand\_4.1.  
 DR InterPro: IPR000048; 10 region.  
 DR InterPro: IPR000857; MYTH4.  
 DR InterPro: IPR001452; SH3.



```

FT DOMAIN 1 95 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 96 114 S1 OF REPEAT I (POTENTIAL).
FT DOMAIN 115 132 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 133 152 S2 OF REPEAT I (POTENTIAL).
FT DOMAIN 153 163 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 164 183 S3 OF REPEAT I (POTENTIAL).
FT DOMAIN 184 187 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 188 206 S4 OF REPEAT I (POTENTIAL).
FT DOMAIN 207 225 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 226 245 S5 OF REPEAT I (POTENTIAL).
FT DOMAIN 246 311 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 312 326 S6 OF REPEAT I (POTENTIAL).
FT DOMAIN 327 357 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 358 483 S1 OF REPEAT II (POTENTIAL).
FT DOMAIN 484 502 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 503 517 S2 OF REPEAT II (POTENTIAL).
FT DOMAIN 518 537 S2 OF REPEAT II (POTENTIAL).
FT TRANSMEM 538 545 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 546 563 S3 OF REPEAT II (POTENTIAL).
FT TRANSMEM 564 574 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 575 593 S4 OF REPEAT II (POTENTIAL).
FT TRANSMEM 594 612 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 613 632 S5 OF REPEAT II (POTENTIAL).
FT TRANSMEM 633 685 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 686 710 S6 OF REPEAT II (POTENTIAL).
FT TRANSMEM 711 1156 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1157 1174 S1 OF REPEAT III (POTENTIAL).
FT TRANSMEM 1175 1190 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1191 1210 S2 OF REPEAT III (POTENTIAL).
FT TRANSMEM 1211 1222 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1223 1241 S3 OF REPEAT III (POTENTIAL).
FT TRANSMEM 1242 1251 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1252 1270 S4 OF REPEAT III (POTENTIAL).
FT TRANSMEM 1271 1289 S4 OF REPEAT III (POTENTIAL).
FT DOMAIN 1290 1309 S5 OF REPEAT III (POTENTIAL).
FT TRANSMEM 1310 1396 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1397 1421 S6 OF REPEAT III (POTENTIAL).
FT TRANSMEM 1422 1476 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1477 1495 S1 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1496 1510 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1511 1530 S2 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1531 1538 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1539 1557 S3 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1558 1566 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1567 1585 S4 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1586 1604 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1605 1624 S5 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1625 1646 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1647 1711 S6 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1712 2339 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 2051 2055 POLY-HIS.
FT DOMAIN 2119 2123 POLY-SER.
FT DOMAIN 2319 2324 POLY-GLY.
FT DOMAIN 379 396 BINDING TO THE BETA SUBUNIT (BY SIMILARITY).
FT NP BIND 452 459 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT SITE 314 314 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT SITE 663 663 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT SITE 1370 1370 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT SITE 1658 1658 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT MOD_HIS 1722 1722 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT CA_BIND 1740 1751 HY SIMILARITY.
FT CARBOHYD 256 256 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 1566 1566 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 1678 1678 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 1678 1678 N-LINKED (GLCNAc. . .) (POTENTIAL).
SO SPOUNCE 2339 AA: 261178 MM: 0413DA93794C8B34 CRC64:

```

Query Match 4.98; Score 101; DB 1; length 2339;  
 Best Local Similarity 21.38; Pred. No. 21;  
 Matches 70; Conservative 34; Mismatches 111; Indels 114; Gaps 15;

```

QY 92 TLTESDSGVYAGAGMNTDRKIOKV-----TLNISEYE-----PSNEQPMPEP 139
DB 1943 TQRTQD-----VLCEAAPIERCHSAFIPVQUGTILAVIVQMGNTLISQIDAPIQOLHS- 1947
QY 140 KWFHLEPYLFQMPAYASSKFEVTRVTPAQRGVPPVHNSPITUIIHPRVSKSSVAGD 199
DB 1998 -----QGRASMPFLAAETQPAD--DASPMKRSTST--LAPFHTRALRGSTALD 2042
QY 200 KPTPIIPSTTASGISALRGLIKQTPSYNHHRLHR-----GRALDYSGSGMR----- 248
DB 2043 PF-----AFSAPIHHHHP--HFEFLFK--FSLEK--SLNACTGAFD 2083
QY 249 -----GVGFHILPITLGLFLLALGLVWRKPAVEPPKALSPAPFLAVPMPI--PSQPP-- 303
DB 2084 STVGPG-----LPIGRG-----PQCKRKRRKRQGRGSGR--KQPSSSSHKHAFY 2129
QY 304 -----GSPRPSQNNIYSCPRRARGADMACTGEAP 334
DB 2130 STIRPAPPEPPYKPSLSHSPTSPTRVDEPPHPPQSSSVHSPLSTSGASTPGRBPQ 2169
QY 335 VP--GPGAPLPAPPAQVSHSWH---APS 359
DB 2190 LPOTPLTRPSVYTKTANSSVHFDGAFS 2218

```

Search completed: November 24, 2002, 06:12:18  
 Job time : 20 secs

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Computeren Ltd.

OM protein protein search, using sw model

Run on: November 24, 2002, 06:09:11 ; Search time 33 seconds

(without alignments)  
2436.102 Million cell updates/sec

Title: us-09-651-150b-2

Perfect score: 2055

Sequence: 1 MBRWIMPIYPIPVSCALP11

HQPAAMHENSUNYINVPA 390

Scoring table: BLAST/OM62

Gapop 10 0, Gapext 0 5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL\_21:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_minc:  
8: sp\_organell:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_protist:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_virus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2047	99.6	390	4	060667
2	1157	56.3	422	11	09D8T1
3	190	9.2	731	6	08SP16
4	186	9.1	534	4	096562
5	184	9.0	758	6	09N247
6	183.5	8.9	532	4	08WV6
7	179	8.7	733	6	08S083
8	175	8.5	455	11	0920L8
9	175	8.5	535	11	09C0T7
10	170.5	8.3	305	11	08VIM2
11	161	7.8	307	11	054947
12	149	7.3	282	11	08VIM1
13	143.5	7.0	299	4	0908K4
14	143.5	7.0	294	4	0908N4
15	129.5	6.3	332	4	08IA55
16	126	6.1	221	11	08VCH2

17	126	6.1	359	4	043656	043656 homo sapien
18	124	6.0	298	4	094097	094097 homo sapien
19	122.5	6.0	301	4	095100	095100 homo sapien
20	118	5.7	820	4	060585	060585 homo sapien
21	117	5.7	897	11	070495	070495 mus musculu
22	117	5.7	1106	4	0810N9	0810N9 homo sapien
23	116.5	5.7	364	4	096042	096042 homo sapien
24	114.5	5.6	335	13	09VGV5	09VGV5 gallus gall
25	113.5	5.5	1417	12	067631	067631 gallus gall
26	113	5.5	1537	5	09VAL2	09VAL2 drosophila
27	112	5.5	2082	16	09S200	09S200 streptomyce
28	111.5	5.4	217	11	08R544	08R544 mus musculu
29	111.5	5.4	392	5	044716	044716 caenorhabdi
30	111	5.4	367	10	09AVC9	09AVC9 oryza sativ
31	111	5.4	801	5	023635	023635 caenorhabdi
32	109.5	5.4	445	14	09PWA4	09PWA4 gallus gall
33	109.5	5.3	862	11	091TK1	091TK1 rattus norv
34	109.5	5.3	2321	12	09D6T6	09D6T6 turkey hezp
35	109	5.3	666	11	099ML2	099ML2 mus musculu
36	108.5	5.3	923	4	09NVH8	09NVH8 homo sapien
37	108	5.3	892	4	09Y438	09Y438 homo sapien
38	107.5	5.2	305	11	08VIM0	08VIM0 mus musculu
39	107	5.2	448	14	09PST1	09PST1 drosophila
40	107	5.2	270	4	09UMT1	09UMT1 homo sapien
41	107	5.2	270	4	09H564	09H564 homo sapien
42	107	5.2	355	2	093R11	093R11 streptococ
43	107	5.2	385	2	054913	054913 streptococ
44	107	5.2	394	10	08RNP3	08RNP3 oryza sativ
45	107	5.2	1275	4	090Q36	090Q36 homo sapien

## ALIGNMENTS

### RESULT 1

060667 PRELIMINARY: PRT: 390 AA.  
 AC 060667;  
 DT 01-AUG-1998 (TREMBLrel. 07, created)  
 DI 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE ANTI-FAS induced apoptosis (Regulator of FAS-induced apoptosis).  
 GN FOS.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 KX MEDLINE:98246048; PubMed:9586636;  
 RA Hiroshi Y., Lorenz J., Kitada S.I., Fisher J., Labarge M., Ring H.Z.,  
 RA Francke U., Reed J.C., Kinoshita S., Nolan G.P.;  
 RT T cells.  
 RT T cells.  
 KI Immunity 8:461-471(1998).  
 KN 121  
 RC SEQUENCE FROM N.A  
 RP TISSUE-B-CELL;  
 RC Straussberg R.;  
 KL Submitted (Aug-2001) to the EMBL/GenBank/DDBJ databases.  
 DR FBL: AF057557; AAC18830.1;  
 DR FBL: H0006401; AAH06401.1;  
 DR InterPro: IPR003599; Iq;  
 DR InterPro: IPR003006; Iq\_MHC.  
 DR Pfam: PF00047; Iq; 1;  
 DR SMART: SMO0409; Iq; 1;  
 SO SPOUNCH: 390 AA, 43146 MW, 553121335A9956 GRC04;  
 Query Match: 99.6%; Score 2047; DB 4; Length 390;  
 Best Local Similarity: 99.7%; Pred. No. 1e-168;  
 Matches 389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Dh	1	MDENIMPELYELPVSAGLRILPEVKEGSELGSSVTKCPDPEMHVRIYLCEMAGSOTGCT	60
QY	61	VSVSTENFIAAFYKGRVTLKQYPRKRLFIYVHTQITFSDSVYACGAGMNTDCKTQKYL	120
Dh	61	VSVSTNFKKFKGRVTLKQYPRKRLFIYEVTQITFSDSVYACGAGMNTDCKTQKYL	120
QY	121	NVISEVEPSSMEQPMETPFKMFHLYELQMPAVYASSKFLVTRTFPAQRKVPVHHSF	180
Dh	121	NVISEVHPSSMEQPMETPFKMFHLYELQMPAVYASSKFLVTRVITPAQRKVPVHHSF	180
QY	181	TTQITHPRRSRASSVATKRPPTPLPSTTAKISALEYLLKPGTYSYHHTRLHPQALD	240
Dh	181	TLVITHPRRSRASSVATKRPPTPLPSTTAKISALEYLLKPGTYSYHHTRLHQRALD	240
QY	241	YGSQSGHEGCGFHLLPTTLGLFLALGLVKKAVEERRKALSRKARHLAVRRKALFSSQ	300
Dh	241	YGSQSGHEYLRHHTLLPTTLGLFLALLVKKAVEPRFALSKARHLAVRRKALESSQ	300
QY	301	PPRASPSPSPSNNIYSAPPRARCAADAAATGEADVPGPGLPPLPADLYSESFWLHATSL	360
Dh	301	KPMQSPSPSPSNNIYSACPRPARCAADAAAGTGEADVPGCAPPLPADLYSESFWLHAPSL	360
QY	361	KTSCEVSYLHQPAAAMEDSDSDQVINPVA390	
Dh	361	KTSCEVSYLHQPAAAMEDSDSDQVINPVA390	

RESULT 2			
Q9DR11			
ID	Q9DR11	PRELIMINARY:	PRT: 422 AA.
AC	Q9DR11		
DT	01-JUN-2001 (TrEMBLref, 17, Created)		
DT	01-JUN-2001 (TrEMBLref, 17, Last sequence update)		
DT	01-JUN-2002 (TrEMBLref, 17, Last annotation update)		
DE	1A10037B05RIK Protein		
GN	1A10037B05RIK		
OS	Mus musculus (Mouse)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
PN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STEPAIN-C57RI/6T; TISSUE-PANCREAS;		
RX	MEDLINE:21085660; PubMed:11217871;		
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,		
RA	Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,		
RA	Aizawa K., Iizuka Y., Nishi K., Kiyosawa H., Kondo S., Yamada K.,		
RA	Saito T., Okazaki Y., Gotohori T., Hono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Glissl C., King H., Koehwa H.,		
RA	Kuchl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,		
RA	Schirrali L.M., Staubli F., Suzuki R., Tomita M., Wanner L., Washio T.,		
RA	Blake K., Okido T., Furuno M., Aono H., Balarelli P., Barsh G.,		
RA	Blauke J., Bottelli D., Bojunga N., Catinici P., de Bonaldo M.F.,		
RA	Browstein M.J., Bull C., Fletcher T., Fujita M., Gariboldi M.,		
RA	Gustineich S., Hill P., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombacres P.,		
RA	Nordone P., Ring B., Schonwald M., Rodriguez I., Sakamoto N.,		
RA	Saeki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-P.,		
RA	Yasuhiki H., Toyouchi K., Wang K.H., Wetzl C., Whitaker C., Wilming L.,		
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuni S.,		
RA	Hayashizaki Y.,		
RT	"Functional annotation of a full-length mouse cDNA collection."		
PI	Nature 409:685-690(2001).		
DP	EMBL: AK007714; EMB25207.1;		
DR	MGD: MG1:1916419; 1810037B05RIK		
DR	InterPro: IPR003600; Iq_1ike		
DR	SMART: SM00410; Iq_1ike; 1;		
SO	SOURCE: 422 AA; 47532 MW; 2597083ASANDREF CPGC4;		

Query Match	56.38;	Score 1157;	DB 11;	Length 4222;
Best Local Similarity	57.58;	Pred. No. 9.6e-92;		

	Matches	234: Conservative	46: Mismatches	107: Indels	20: Gaps
QY	1	MDKRLMLTYLTYPLVSGARILRLPKYKVGCHLGGSTVYLKCPILPKMKHRYVILCRKMMASGTCGT	60		
Db	1	MDKRLMLTYLTYPLVSGARILRYEVLQVNVHSDSILTECPILQVHVMYILCRQMAAPGICST	60		
QY	61	VSTTFNFKAEYKGRVTLKQYPRKNIPLTEVQVLESISGVAYASVACSMNIDKCKTOKVTL	120		
Db	61	VYSTVFKKKEFYERRVTLPTGCDLKLILVYMLQVLPENDDGLTACQGVCMKIDKCKTOKITL	119		
QY	121	NVHSEYFIPSEBQPMPTPKWEHLTYLFLQMP-----AYASSKRYVLTTPAQCCKVPP	174		
Db	120	NVHSEYEPPEWDEDMYTSERPRMLHRELIQHMPLWLSGHPSSSVGLAKVTTPKTEAPV	179		
QY	175	VHNSPTTQGLTHPRVSPASVAGCKPRLPLSTTASKISALAEGLKCKQTPSYNHNHRLH	234		
Db	180	VHQSSTTSTVQIHRPKYKRAVSATKSPALPLATTASKISTQQA-IKRLASVSHNHRH	238		
QY	235	QVQALTYVSGSGREDSQG-----FHTLPTVLTFLLALSLVVKRPAVERPKALSPRAR	297		
Db	239	EQRTLRHHCPRYGRHDKQILNLPPEPHTLIPPTGCLIVLIGLVKRAICQVPAASSPAP	298		
QY	288	KLAVRMALHSS-----QVRKSSPKKRSQNNIYSACPRKAKCAADAGICGAVGCPAPL	342		
Db	299	FLAMPKPGPGASPPPTQRPDASQPRSSQNNYSQMPAPKAPQSLPFAEAPILLNAPASA	358		
QY	343	PPATIGVSPHMLHAPSLKTCSEYVSLYQQAPAMHNDISDDYINVP	389		
Db	359	SPASQVYLKAPVPHHTSLAKSCSYVSLCYQVAVNLFDVDSIDYINLP	405		

```

RESULT 3
OSBSP16
LU OSBSP16 PRELIMINARY; PRT: 731 AA.
AC OSBSP16
DT 01-JUN-2002 (TrEMBLref. 21, Created)
DT 01-JUN-2002 (TrEMBLref. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLref. 21, Last annotation update)
DE Polymetric immunoglobulin receptor.
FE PIGR.
OS Macropus eugenii (Tamar wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metabacteria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9315;
KN [1]
KN RP SEQUENCE FROM N.A.
KN RC TISSUE=MAMMARY LYMPH NODE;
RA Taylor C.L., Harrison G.A.;
RT *cDNA cloning of the polymetric immunoglobulin receptor of the
RT marsupial Macropus eugenii (tamar wallaby).";
RL Submitted (OCT-2000) to the EMBL/Genbank/DDIJ databases.
DR EMBL; AFJ17205; AAK69593.1; -.
KM Receptor.
SQ SEQUENCE 731 AA: 80253 MW: 91467360.34786017 CRC64;

```

[illegible]

## RESULT 4

Q96SA2 PRELIMINARY: PRT: 514 AA.  
 ID Q96SA2: AC Q96SA2: RT  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE FKS687 protein.  
 GN FKS687.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang Y.-G., Comp 1;  
 RT "Molecular cloning and characterization of FKS687, a novel gene  
 RT located on human chromosome 1.";  
 RL Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF354295; AAK39522.1;  
 SV SEQUENCE 514 AA; 56748 MW; 6EF8050E412AF91C CRC64;

Query Match 9.18; Score 186; DB 4; Length 514;  
 Best Local Similarity 23.28; Pred. No. 1.1e-07;  
 Matches 98; Conservative 57; Mismatches 151; Indels 136; Gaps 20;

QY 3 RMLPFLFLPVSGALRIIPVYK-----VEGLGGSVTIKCP 38  
 DB 45 RMLPFLFLPVSGALRIIPVYK-----VEGLGGSVTIKCP 103  
 QY 39 LP---EMHVRITLCREMAAGSGTGVSTNFKAEKKGAVTLKYPKKNLFLVEYTL 94  
 DB 104 YAPSSVNHQKRYKWRRLAPPWITQTIVSTNQTHHPRVATLTPPQGLVVRISQL 163  
 QY 95 TESUSGVYACGAGNTDQKTKOKVTLNVHSHYSPSWEPQMP--ETPKWPHIPLYFQMPAY 153  
 DB 164 SPDDIGCYLCIG--SENMLFLSMNLITISGPA---STLPATIPAAAGEL---TMRSY 213  
 QY 154 ASSKFTVTRVITPAQKGVPPVHHSSPTQTITPRVSRASSVAGDK--PRTLPSTTAS 211  
 DB 214 GTASPVANKWTP---GTTQTLCOGTAMDIVASTPGTSTTASAGCRRTGATIRPAAPGT 269  
 QY 212 KISALEGLK-----POTPSYNNHITRLHROKALDYSQSGKEGGPILLITLIGLELLA 266  
 DB 270 G-SMAEVSVKAPAPIPESPSPKSKMSNTTEGWEGRSS----- 308  
 QY 267 LIGLVVAKA---VKKRAKLSRRAR---LAVMRALFESSOR----- 301  
 DB 309 ---VTNRAPASKDRREMTTKADPPREDIEGVPI-ALDAAKVLTGTPPALVSETLAW 363  
 QY 302 ---PRGSP--PPRSQNNIYSMPPRAK;ATAAGTGEAPV----- 335  
 DB 364 KLIQATVHVSQSGSGTETTP--AAGMWTLGTADVWMTSMFAASGPGSAAGLDAA 421  
 QY 336 ---PGGAPLPPAPLOVSESPW 354  
 DB 422 TGDGPGQATLSQTP---AVGPPW 440

RESULT 5  
 Q9N2H7 PRELIMINARY: PRT: 758 AA.  
 ID Q9N2H7: AC Q9N2H7: RT  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Poly-Ig receptor precursor  
 OS Sus scrofa (Pig).  
 GN Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactylia; Suina; Suidae; Sus.  
 NX NCBI\_TaxID=9923;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stone T., Kumura H.;

"Porcine mammary gland cDNA clone, similar to poly-Ig receptor."  
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB032195; AAA84283.2;  
 DR InterPro: IPR003599; 19;  
 DR InterPro: IPR003600; 19\_1like;  
 DR InterPro: IPR030066; 19\_1like;  
 DR Pfam: PF00047; 19; 5  
 DR SMART: SM00409; 19; 3;  
 DR SMART: SM00410; 19\_1like; 2.  
 KM Receptor; Signal.  
 FT SIGNAL 1  
 SO SEQUENCE 758 AA; 83154 MW; D5BED1AB082D247 CRC64;

Query Match 9.08; Score 184; DB 6; Length 758;  
 Best Local Similarity 34.28; Pred. No. 2.6e-07;  
 Matches 41; Conservative 22; Mismatches 41; Indels 16; Gaps 4;

QY 17 LRIPEVEVEL-----GSSVTKPLP---EMHVRITLCREMAAGSGTGVSTNFKAEKKGAVTLKYPKKNLFLVEYTL 62  
 DB 10 LAIFPVVSMKSHIFGQDVSSVSSVSIFRYPATSVNHRSEKWF--IGAKRPTTLI 68  
 QY 63 STTNFKAPVKGQVITKQYPPKKNLFLVEYTLTESUSGVYACGAGNTDQKTKOKVTLNVHSHYSPSWEPQMP--ETPKWPHIPLYFQMPAY 153  
 DB 69 SSEGVTSKQYKGFANLTFENGIIFVMTIGHLTQSTSLYKQSGSGSS--FGLSEFVLEV 127

## RESULT 6

Q8MWV6 PRELIMINARY: PRT: 532 AA.

ID Q8MWV6: AC Q8MWV6: RT  
 DT 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Fc alpha/mu receptor.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21638011; PubMed=11779189;  
 RT "Expression of Fc alpha/mu Receptor by Human Mesangial Cells: A  
 RT Candidate Receptor for Immune Complex Deposition in IgA Nephropathy."  
 RL Hiochem. Biophys. Res. Commun. 290:438-442(2002).  
 DR EMBL: AY063125; AAL51154.1;  
 DR InterPro: IPR003599; 19;  
 DR SMART: SM00409; 19; 1.  
 KM Receptor.  
 SV SEQUENCE 532 AA; 57143 MW; D347A23C0FA1EED3 CRC64;

Query Match 8.98; Score 183.5; DB 4; Length 532;  
 Best Local Similarity 23.08; Pred. No. 1.8e-07;  
 Matches 93; Conservative 56; Mismatches 147; Indels 109; Gaps 18;

QY 3 RMLPFLFLPVSGALRIIPVYK-----VEGLGGSVTIKCP 38  
 DB 25 RMLPFLFLPVSGALRIIPVYK-----VEGLGGSVTIKCP 83  
 QY 39 LP---EMHVRITLCREMAAGSGTGVSTNFKAEKKGAVTLKYPKKNLFLVEYTL 94  
 DB 84 YAPSSVNHQKRYKWRRLAPPWITQTIVSTNQTHHPRVATLTPPQGLVVRISQL 143  
 QY 95 TESUSGVYACGAGNTDQKTKOKVTLNVHSHYSPSWEPQMP--ETPKWPHIPLYFQMPAY 153  
 DB 144 SPDDIGCYLCIG--SENMLFLSMNLITISGPA---STLPATIPAAAGEL---TMRSY 193  
 QY 154 ASSKFTVTRVITPAQKGVPPVHHSSPTQTITPRVSRASSVAGDK--PRTLPSTTAS 211  
 DB 194 GTASPVANKWTP---GTTQTLCOGTAMDIVASTPGTSTTASAGCRRTGATIRPAAPGT 249  
 QY 212 KISALEGLK-----POTPSYNNHITRLHROKALDYSQSGKEGGPILLITLIGLELLA 266

DB 250 G-SWAGSVKAPAPIPESPSPSKRSMSNTTEGWEGRSS----- 288  
 QY 267 LAGLVVKKK-----VKKKALSKRAK-----LAVKKALPSSOR----- 301  
 DB 289 ----VTKRASKDRKEMTTTKADKRRHDEVKRI-ALDAAKKVICTIGPAIVSEPIIAW 343  
 QY 302 ---PRGSP---RPSQNNIYSACPRRARRGADAGTEAPVPGAP 341  
 DB 344 ELIPQATPVSKODSOSIGETTP--AACWMLIGTPAIVWILGTP 386

## RESULT 7

Q8S083 PRELIMINARY: PRT: 733 AA.

ID Q8S083  
 AC Q8S083  
 DT 01-JUN-2002 (TREMBlrel: 21, Created)  
 DT 01-JUN-2002 (TREMBlrel: 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel: 21, Last annotation update)  
 DE Polymeric immunoglobulin receptor.  
 GN PIGR.  
 OS Trichosurus vulpecula (brush-tailed possum).  
 OC Mammalia; Metazoa; Chordata; Vertebrata; Eutelestomi;  
 OC Mammalia; Metazoa; Chordata; Vertebrata; Eutelestomi;  
 OX MBL\_TaxID-9337;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 KC TISSUE=MAMMARY GLAND;  
 RA Adamski F M., Demmer J.;  
 RT Cloning and characterization of PIGR and J chain of the marsupial,  
 RT Trichosurus vulpecula (brush-tailed possum).  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF091137; AADA1688.1;  
 KW Receptor.  
 SW SH02CHF 733 AA; 80300 MW; 81688720BAC14A5A CRC64;

Query Match 8 7%; Score 179; DB 6; Length 733;  
 Best Local Similarity 35.0%; Pred. No. 6, 7e-07;  
 Matches 42; Conservative 21; Mismatches 41; Indels 16; Gaps 4.

QY 17 LRLIPV-----KVEHLGSGVTKCHLP-----EMHVRIVLCREMGSGCICGV 62  
 DB 9 LALPVSMSKSPFPEKQVTVGERSVSIOCFYSSVNPVHKKYF\*LQNLPG-SCEITV 57  
 QY 63 STNFKAKYKGPVITKQYPPKNLFLVEVTLTESGVSVMGGMNDRECKTKVILNV 122  
 DB 68 SSNGVSEPSRSPKRTINPPGNNSPIQLISLEKEDJGLYKGLG-ITNKGISPDITLEV 126

## RESULT 8

Q92018 PRELIMINARY: PRT: 455 AA.

ID Q92018  
 AC Q92018  
 DT 01-DEC-2001 (TREMBlrel: 19, Created)  
 DT 01-DEC-2001 (TREMBlrel: 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel: 20, Last annotation update)  
 DE Fca/m receptor (Fragment).  
 GN FCAMR.  
 OS Mus musculus (Mouse).  
 OC Fukuayota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;  
 OC Mammalia; Eutelestomi; Rodentia; Sciurognathi; Muridae; Mus;  
 OX MBL\_TaxID-10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 KC Shimizu Y., Honda S., Yotsumoto K., Tahara Hanaka S., Eyre H.J.,  
 KA Sutherland G.P., Fudo Y., Shibuya K., Koyama A., Nakachi H.,  
 RA Shibuya A.;  
 RT \*Fca/m receptor is a single gene-family member closely related to  
 RT polymeric immunoglobulin receptor on chromosome 1.\*;  
 RL Immunogenetics 0:0-0(2001)  
 DR EMBL: AH071978; BAB71750.1;  
 DR InterPro: IPR003006; Iq\_MHC.  
 DR InterPro: IPR003596; Iq\_V.  
 DR InterPro: IPR00508; Sigptase.

DR Ffam, FFG0047, 19, 1.  
 DR SMART: SM00406; IGV: 1.  
 DR PROSITE: PS00501; SPASE\_1.1; UNKNOWN\_1.  
 KW Receptor.  
 FT NON-ITER 1  
 SQ SEQUENCE 455 AA; 48810 MW; 628913034A4C365 CRC64;

Query Match 8.5%; Score 175; DB 11; Length 455;  
 Best Local Similarity 23.5%; Pred. No. 8, 1e-07;  
 Matches 89; Conservative 39; Mismatches 136; Indels 114; Gaps 14;

QY 25 VEEIAGSVITKPLP-----EMHVRIVLCREMGSGCICGV 60  
 DB 10 VTSNTSGAVTTHIVAPSSVNRHCKYKWLKSSPLMTIHTVASTNQTHIDYPSALTD 69  
 QY 81 YPKNKLIVEVQULHESQVYACACANNITQKQKQKTLNVHSHYRPS---WEEQMPH 137  
 DB 70 VPOSILVAVRLPLSLDGLVGLKPSH---LPSNMLFESVNLIVSNPSTYAMAPASS 126  
 QY 138 TPKWHLPYLFQMPAYASSSKFEVITTPAGKGVPEVHSHSPITQITHPFVFASSVA 147  
 DB 127 KIP-----FVASPGASS-----ACNG-----WISGVQILHG-----S 154  
 QY 198 GKKPPTFLPSTASKISALESLKQTPSYNHITLHPGALDVSQSGREGQGHILLIP 257  
 DB 155 GSEWRTAPPTGTSTK-----TTSNANPQTLPTAPTVIITGSPREGS----- 197  
 QY 258 TIIQIPLALLGIVKRAVERKKALSRARHIAVMR-----ALPSSQPRK 304  
 DB 198 -----IRAAVPTPEGSPKSKSSSTTQVWLMNTNSVTSVITSEGRQ 243  
 QY 305 SPPRQNNIYSACPRRARRGADAGTEAPVPGAPLPAPL-----QVSE 351  
 DB 244 GTTPPTTG-----PRDE---TIVRVSPPAPKKTCTTPSPALISMHVWITIGKTRVSK 295  
 QY 352 SFWLH-----AFSIKT 362  
 DB 296 VQMLSLLELSPASPAV 313

## RESULT 9

Q9E0T7 PRELIMINARY: PRT: 535 AA.

ID Q9E0T7  
 AC Q9E0T7  
 DT 01-MAR-2001 (TREMBlrel: 16, Created)  
 DT 01-MAR-2001 (TREMBlrel: 16, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel: 20, Last annotation update)  
 DE Fca/m receptor.  
 GN FCAMR.  
 OS Mus musculus (Mouse).  
 OC Fukuayota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;  
 OC Mammalia; Eutelestomi; Rodentia; Sciurognathi; Muridae; Mus;  
 OX MBL\_TaxID-10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 KC MEDLINE 21170225; PubMed 11062505;  
 RA Shibuya A., Sakamoto N., Shimizu Y., Shibuya K., Osawa M.,  
 KA Hironaka T., Eyre H.J., Sutherland G.P., Fudo Y., Fujita T.,  
 KA Miyabayashi I., Sakano S., Tsuji T., Nakayama K., Phillips J.H.,  
 KA Lanier L.L., Nakachi H.;  
 RT \*Fca/m receptor mediates endocytosis of IgM-coated microbe.\*;  
 RL Nat. Immunol. 1:441-446(2000).  
 DR EMBL: AH048834; BAB7312.1;  
 DR MGD: MGI:1927803; Fcamr.  
 DR InterPro: IPR003599; Iq\_V.  
 DR InterPro: IPR003600; Iq\_Like.  
 DR InterPro: IPR003006; Iq\_MHC.  
 DR InterPro: IPR00508; Sigptase.  
 DR Pfam: PF00047; Iq\_1.  
 DR SMART: SM00409; Iq\_1.  
 DR SMART: SM00410; Iq\_Like; 1.  
 DR PROSITE: PS00501; SPASE\_1.1; UNKNOWN\_1.  
 KW Receptor.

SQ SEQUENCE: 535 AA; 57696 MW; H275R77C70151C75 CRG64;

Query Match	8.58	Score 175;	DR 11;	Length 535;
Best Local Similarity	23.58;	Pred No 9	92-07;	
Matches	89;	Conservative	39;	Mismatches 136;
				Indels 114;
				Gaps 14;

```

QY 25 VGGGIGGSAVIGKPIR-----EMHVRVLYGEMAGSGTGVIVSTINIKKAVKQVNTLKQ 80
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 90 VTGNTGTAATITPHTAPRSVNRKLEKTVKZLZSLMPLHTGVSTNLTHTHPRVGRFALLTP 149
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 81 YPKKNILFIVEVTLQTESGNYUATGAGMNDPRKTKVTLNVHSYERS---WEEDPRE 137
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 150 VPQSGFVAVLLRLSLGVGVGKRCGIDG---DRNDMLFVSFVNLTVSAGPSNTTVAAAIASS 206
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 138 TPKWFLRFLQMVAVVASSKFFVTRVTPAKKQKVVYVHNSSEFTLITHRFVAVASSVA 197
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 207 EP-----TTSPPVANS-----AGNG-----WTSVTLQED-----S 244
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 198 GQKPKPLPSTTAKSISALPGLKKQTPSYNNHNPRIHQPRALDYGSGQKQKQCPHHP 257
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 245 GSEWIGTAAFTTGTSTR-----TTSANPPLQTLFATVAVLDTGSPSEBS----- 277
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 258 TILGLLLALLDLVKKVAVERRPKASPRAPPLAVPAP-----ALESSQRPBG 304
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 278 -----IRAAVPTPECPSPKSKSMSTIIQGVAMLNTHNSVTPVSTHSQRKQ 323
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 305 SPRPSQNNIVYSAPPRAPRADAAGTDEADVPPGAPRLPAPU-----GVSE 351
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 324 GTPELIDG-----PRDE--TDVAVSPEDARKTTGTTRESALISEHWETWELLQKTEYSK 375
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 352 SPWLIH-----APSLIKT 362
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 376 QOMLHSLLEELSPASQIT 393
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESID: 10
QYVIM2
DbVIM2
AC QYVIM2: PPT:IMINAVY: PRT: 305 AA.
DT 01-MAR-2002 (TREMDELrel. 20, Created)
DT 01-MAR-2002 (TREMDELrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMDELrel. 21, Last annotation update)

```

```

GENEID 10
OBSID2
ID OBSID2: PRIMARY: PRT: 305 AA.
DT 01-MAR-2002 (TREMblrel, 20, Created)
DT 01-MAR-2002 (TREMblrel, 20, Last sequence update)
DT 01-JUN-2002 (TREMblrel, 21, Last annotation update)
DE TM1.
GN TM1d.
OS Mus musculus (Mouse).
OC Eukaryota, Eumetazoa, Chordata, Craniata: Vertebrata: Euteleostomi;
OC Mammalia, Eutheria, Rodentia: Sciurognathi, Muridae, Murinae; Mus
OX NCBI_Taxid=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-SPLEEN;
RA McMillen J.C., Imelso S.E., Akbari O., Potter M., Barsh G.S.,
RA Freeman G.T., Imelso D.T., Dekryll R.H.;
RT Tmap, a major T cell regulatory locus that controls the development
R1 of airway hyperactivity, cosegregates with variants in a novel gene
R2 family.
RL Submitted (JUL-2001) to the EMBL/GenBank/DBD databases.
DR EMBL: AF956829; AAL35774.1.
DR MGD: MG1:2159680; Tm1d1.
DR InterPro: IPR003599; Tg.
DR InterPro: IPR003066; Tg_MHC
DR Pfam: PF00047; Tg: 1.
DR SMART: SM00409; Tg: 1.
SU SEQUENCE 405 AA: 3391 MW: 8F4EA36627FE65FB C70F64:

```

Query Match	8.38;	Score 170.5;	DB 11;	Length 305;
Best Local Similarity	24.48;	pred. NO. 1.2e-06;		
Matches 84, Conservative	49;	Mismatches 118;	Indels 93;	Gaps 18

DQ 13 VSGALRLP-----EVRVEGELSSSVTKPLPEMHVRIYL--KEMASSTQTGVAST 65  
          ::: ::         | : | : | : | : | :  
Db 8 ISGLILLPGIVDSYVFVKGVGHPTLPCTYSTYRCITTTTCMGRCQCPSAQQNTLLWT 67

QY	66	NPKAVYK--GAVTI KQVKKNI FLVVTGI IISGVCAAGCA---GNNIIIPKQIGVLI	120
Db	68	NHPIVTVYQASPPRYNIKRGHSESTVSLITENSSESTSLCTGVLEIPAFND---GKVF	123
QY	121	NHSEVEYPMEEQPMPELPKMFHLPYLCMPAYASSKKEVTVIT--PAQPK-----	171
Db	124	SI-----QVMPKIPD-----IPDIPPIIIIPPAIGSHPIISIK	156
QY	172	---VPP---VHHSFTQL---THKPPVSSASSVAQKIPLELEFTISKSL-----	216
Db	157	STHPTSPISVSTSTPTPTSTHTWTKP-----EPITCPMETTAEVGTGFSHTPT	205
QY	217	--GCLLKPDTPVNNHTI HQPAILDVGSGSGHPGQPHILPTIIIGLILGLGVKK	274
Db	206	DWNVTIASSDWNINNT-----EAIPEKQKQKPTGSGFVGI--GVALLLLLLVSTV---	256
QY	275	AVPRKALSPRAPPIAV---PMRLSSQPPPSPPPKSNNYI	315
Db	257	AIIPYIIIMKRSASLSVYAFVPSKIFALQN-AAVHSHAPENIY	299

RESULT 11

054947

ID

054947

PRELIMINARY

PRI

307

AA

AC

054947

DT

01-JUN-1998 (TREMblrel)

06

Created

DT

01-JUN-1998 (TREMblrel)

06

Last sequence update

DT

01-JUN-2001 (TREMblrel)

17

Last annotation update

DE

Kidney injury molecule-1 precursor (KIM-1).

GN

KIM-1.

OS

Rattus norvegicus (rat).

OC

Eukaryota, Metazoa, Chordata, Vertebrata, Euteleostomi.

OC

Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Rattus.

OX

NCBI\_TaxID=10116;

RA

NCBI\_TaxID=10116;

RP

SEQUENCE FROM N A

RC

STRAIN=STRACU-DAMLEY; TISSUE=KIDNEY;

RX

MEDLINE=98129827; PubMed=9461608;

RA

Ichimura T, Benveniste J V, Bailly V, Wei H, Hession C A, Gate R L, Santoluc M.

RT

\*Kidney injury molecule-1 (KIM-1), a putative epithelial cell adhesion molecule containing a novel immunoglobulin domain, is up-regulated in renal cells after injury \*.

RL

J. Biol. Chem. 273:4135-4142(1998).

CC

-1- FUNCTION: MAY ACT AS AN EPITHELIAL CELL ADHESION MOLECULE.

CC

-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).

CC

-1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN LIVER, SPLEEN AND NORMAL KIDNEY. LEVELS INCREASE IN THE POSTISCHEMIC KIDNEY WITH EXPRESSION FOUND IN REGENERATING PROXIMAL TUBULE EPITHELIAL CELLS

CC

-1- INDUCTION: IN RENAL CELLS, AFTER INJURY

CC

-1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS ONE V-LIKE DOMAIN.

CC

EMBL: AF035963; AAC35946.1; -

DR

InterPro: IPR003599; 19.

DR

InterPro: IPR003096; 19\_MHC.

DR

Pfam: PF00047; Ig: 1.

DR

SMART: SM00409; Ig: 1.

KW

Transmembrane; cell adhesion; glycoprotein; signal.

FT

SIGNAL

1

21

POTENTIAL.

FT

CHAIN

22

307

KIDNEY INJURY MOLECULE-1.

FT

DOMAIN

22

235

EXTRACELLULAR (POTENTIAL).

FT

TRANSHEM

236

256

POTENTIAL.

FT

DOMAIN

257

307

CYTOPLASMIC (POTENTIAL).

FT

DOMAIN

23

130

IG-LIKE V-TYPE DOMAIN.

FT

DOMAIN

33

37

SIGNATURE FROM IGS CONSTANT REGIONS.

FT

DOMAIN

131

201

THR/SER/PRO RICH.

FT

DOMAIN

131

234

MUCIN-LIKE.

FT

DOMAIN

138

152

3 X 5 AA TANDEN REPEATS OF R-P-I-I-I-I-I.

FT

REPEAT

138

142

1-1.

FT

REPEAT

143

147

1-2.

FT

REPEAT

148

152

1-3.

FT

DISULFID

37

108

POTENTIAL.

FT

CARBOHYD

206

206

N-LINKED (GLICNAAC... ) (POTENTIAL).

FT CARBOHYD 218 218 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 307 AA: 33966 MW: 746016DID1549/60 CMO64;  
 Query Match  
 Host local Similarity 25.1%; Pred. No. 7.8e-06;  
 Matches 89, Conservative 54, Mismatches 119, Indels 94, Gaps 28;  
 QY 13 VSGALRLP-----EKKVELESGSVTKRPLPEMHVPLV---REMAVSTGTAVST 64  
 DB 8 ISGLILLIGADVSTVY-VKGVGHVPIIPCTYSTRCGIIITTCWKGCCPSSCONILLW 66  
 QY 65 TNEIKAEYK--GRVTLKQYPRKNLFLVEVTLQTESGSAVYAG--GMNTDREGTKVT 119  
 DB 67 TNGVQVTPSSSPYNIKIPISGVSLTIENVEDSDSLYCPVPEIPWFNI---QMT 122  
 QY 120 LNVHSYEPSPWEPMPHPKPHLYLPOMAVASSSKPVTRVTPAQKQVP---V 175  
 DB 123 FSL--EVKPEITSP-PIRP-----TTPTPTPTTISTRTVPTSTRV 165  
 QY 176 HNSPT---TJTHREPRVSSVAGDKRIFLPSTTASKEGLKPLQVPSY---- 247  
 DB 166 STSTPTPTPTQ--THKEPT-----TPVAHETIAVY-----PTPTTADW 205  
 QY 228 -----NHHTRLRQALDYSGSPREGPHILPTLGLFLALLGLVYKPA 275  
 DB 206 NGVTSSEAWNNHTVPIPLK-----PGRNPTKGFYGVMSVALLLLASTVTVTRY 259  
 QY 276 VERRK---AISRRARLAVRNALLESQPRGSPRPSRONNYSACPRARACD 326  
 DB 260 IIRKMGSLSTEVAFIVS-KSRALQNA-----AIVHFRADNITY-ITFRSRQAE 307  
 RESULT 12  
 OBYIMI  
 ID Q9VIM1 PRELIMINARY: PRI: 282 AA.  
 AC Q9VIM1:  
 DT 01-MAR-2002 (TREMblrel) 20, (created)  
 DT 01-MAR-2002 (TREMblrel) 20, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel) 21, Last annotation update)  
 DE 11M1.  
 GN 11M1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus  
 OX NCBI\_TaxId=10090;  
 RN 11  
 RP SOURCE FROM N.A.  
 RC STRAIN-DBA/2; TISSUE-SPLEEN;  
 RA McInliffe J T, Imetisu S F, Akbari O, Potter M, Rash G S,  
 RA Freeman G T, Imetisu D T, Dekruijt R H.;  
 RT "Tapi, a major T cell regulatory locus that controls the development  
 RT of airway hyperactivity, cosegregates with variants in a novel gene  
 RT family.";  
 PL Submitted (TM-2001) to the EMBL/GenBank/DBJ databases  
 PR FMRB: AF349830; AAI35775.1;  
 DR MGD: MGI:2159680; Tlmd1.  
 DR InterPro: IPR003599; Iq;  
 DR InterPro: IPR003006; Iq\_MHC.  
 DR Pfam: PF00047; Iq.1.  
 DR SMART: SM00409; Iq.1.  
 SQ SEQUENCE 282 AA: 30968 MW: 7030EE0698F0ACFE CRC64;  
 Query Match  
 Host local Similarity 7.3%; Score 149; DB 1; Length 282;  
 Matches 77, Conservative 44, Mismatches 132, Indels 66, Gaps 14;  
 QY 13 VSGALRLP-----EKKVELESGSVTKRPLPEMHVPLV---REMAVSTGTAVST 65  
 DB 8 ISGLILLIGADVSTVY-VKGVGHVPIIPCTYSTRCGIIITTCWKGCCPSSCONILLW 67  
 QY 66 NPIKAEYK--GRVTLKQYPRKNLFLVEVTLQTESGSAVYAG--GMNTDREGTKVT 120  
 DB 68 NGRVTVYQKSSRYNLKGHSDEGVSLLTIENVEDSDSLYCPVPEIPWFNI---QKVT 123

QY 121 NVHSYEPSPWEPMPHPKPHLYLPOMAVASSSKPVTRVTPAQKQVP-PVHHS 179  
 DB 124 SL-----QVKPLP-----TPPRKPTPTTAPQK 149  
 QY 180 FTJTHREPRVSSVAGDKRIFLPSTTASKEGLKPLQVPSY---NHHTRLRQAL 239  
 DB 150 PTISTRTVHVTSTRVSTPTPT--STHTMIRKPDWMTISSGDRSNIT-----EAI 202  
 QY 240 DWGSDSGRPGDPRHIIIPILQILALALGVKAVPRKASPRAPVAV---PRAL 296  
 DB 203 PRGPKQNPNTKGFYVGI-QIALLLLLLVSTV---AIFPYILMKRSASLVAVFVSKI 258  
 QY 297 ESSGPFPSFPFPPSONITY 315  
 DB 259 EALQN AAVVHSRADNITY 276  
 RESULT 13  
 OYUBKA  
 ID Q9URK4 PRELIMINARY: PRI: 299 AA.  
 AC Q9URK4:  
 DT 01-MAY-2000 (TREMblrel) 13, (created)  
 DT 01-MAY-2000 (TREMblrel) 13, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel) 19, Last annotation update)  
 DE 1RC1.  
 GN 1RC1A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catartida; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN 11  
 RP SOURCE FROM N.A.  
 RC TISSUE-LYMPHOID;  
 RA O'Connor C.D.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 EN 12  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LYMPHOID;  
 RA Cantoni C., Biassoni R.;  
 RT "Molecular cloning of a novel inhibitory receptor.";  
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN 13  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LYMPHOID;  
 RA Cantoni C., Biassoni R.;  
 RT "JRC1 isoforms.";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 PR FMRB: AF234864; CAB55347.1;  
 DR InterPro: IPR003599; Iq;  
 DR InterPro: IPR003006; Iq\_MHC.  
 DR Pfam: PF00047; Iq.1.  
 DR SMART: SM00409; Iq.1.  
 SQ SEQUENCE 299 AA: 33201 MW: 978461DA87EM6269 CRC64;  
 Query Match  
 Host local Similarity 21.9%; Pred. No. 0.00025;  
 Matches 89, Conservative 48, Mismatches 128, Indels 141, Gaps 20;  
 QY 4 WL-WPLYLPLVPSALPLIPEVKELESGSVTKRPLPEMHVPLV---REMAVSTGT 60  
 DB 2 WLDWALLLWVHOCALSKCKIVAGVGSLSVQGVYHKKRHLINKWCKP-IVPTLQDK 60  
 QY 61 VSTINFAIKAKGVTLKQYPRKNLFLVEVTLQTESGSAVYAGGMNTDREGTKVT 120  
 DB 61 IYETKG-SAGKKNKGVSLIDKSPANISFTVTLLENLDEEAGYKMGV----- 105  
 QY 121 NVHSYEPSPWEPMPHPKPHLYLPOMAVASSSKPVTRVTPAQKQVPVPH 176  
 DB 106 -----DTP-WLQDFHDPVVEVEVSPFASISM-----TAS-----IT 137  
 QY 177 HNSPTTJTHREPRVSSVAGDKRIFLPSTTASKEGLKPLQVPSY---NHHTRLRQ 236

```

Db 138 AAKSTITTAFFPPYS-----STLLFAVGA-----THSASIOEE 170
QY 237 KALIDYGSQSGKGGPHLLIPITLIGLPLALIGI VYKRAVERKALSKRARRIAYRM--- 293
Db 171 TEEVYNSQ-----LPLSLTALLLLLV-----CASLAWMMFOK 206
QY 294 --PALESSOPRPGSPRPSGNNIYSA-----CPPARAGADAMTGEAVVPW:GALPLPA 345
Db 207 WIKAGDHSEI:SONPKQATOSPIHYANI EILMPPIQFK-----PAPPR 249
QY 346 PLOYSESFWLHAPSILKTSCEVSLY-----HOPANM---EDSDSD 383
Db 250 EVEVEYST---VASPREELHYASVPEGSNTNPIAAQFPREEDPSD 292

RESULT 14
QY00GN4 ID QY00GN4 PRELIMINARY: PRT: 299 AA.
AC QY00GN4:
D1 01-MAY-2000 (TIPMBLrel 13, Created)
D1 01-MAY-2000 (TIPMBLrel 13, last sequence update)
D1 01-DEC-2001 (TIPMBLrel 19, last annotation update)
DE NK inhibitory receptor.
GN IR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LYMPHOID;
RA Cantoni C., Hiassoni R.;
R1 IRP60 a novel NK inhibitory receptor.
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ388323; CAB66145.1;
DR InterPro: IPR008599; I9_
DR InterPro: IPR008006; I9_MHC;
DR Pfam: PF00047; I9;
DR SMART: SM00409; I9;
KW RECEPTOR.
SQ SEQUENCE 299 AA: 33172 MW: 3992BDID85C3B63DB CRG64.

Query Match 7 0% Score 143 5, DB 4, Length 299;
Best local Similarity 21.9% Prod. No. 0.00025;
Matches 89; Conservative 48; Mismatches 128; Indels 141; Gaps 20;

```

```

Db 250 EVEVEYST---VASPREELHYASVPEGSNTNPIAAQFPREEDPSD 292

RESULT 15
QY08TA95 ID QY08TA95 PRELIMINARY: PRT: 332 AA.
AC QY08TA95:
D1 01-JUN-2002 (TIPMBLrel 21, Created)
D1 01-JUN-2002 (TIPMBLrel 21, last sequence update)
D1 01-JUN-2002 (TIPMBLrel 21, last annotation update)
DE Similar to CMR35 antigen precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC025395; AAH25395.1;
SQ SEQUENCE 332 AA: 36059 MW: 1591246.13630C93 CRG64;

Query Match 6.7% Score 129.5; DB 4, Length 332;
Best local Similarity 24.6% Prod. No. 0.0046;
Matches 83; Conservative 44; Mismatches 139; Indels 71; Gaps 17;

```

Search completed: November 24, 2002, 06:13:00  
Job time : 36 secs

1911